

AUDIOVISUAL DATA

DATA-DRIVEN PERSPECTIVES FOR MEDIA STUDIES

EDITED BY
GIOGIO AVEZZÙ
MARTA ROCCHI





13TH MEDIA MUTATIONS INTERNATIONAL CONFERENCE

AUDIOVISUAL DATA

DATA-DRIVEN PERSPECTIVES FOR MEDIA STUDIES

EDITED BY

GIORGIO AVEZZÙ

MARTA ROCCHI

First published 2023 by Media Mutations Publishing
via Barberia 4, Bologna, Italy
ISBN 9788894731804
DOI <https://doi.org/10.21428/93b7ef64.ddf0a578>



© The Editor(s) (if applicable) and The Author(s). This book is an open access publication. Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material.

INDEX

1. A Doubtful Handshake: The Promising Frontier of Data-Driven Film and Media Studies	7
Giorgio Avezzù and Marta Rocchi	
2. Exploring TV Seriality and Television Studies through Data-Driven Approaches	23
Mirko Degli Esposti and Guglielmo Pescatore	
3. Looking for Lexical Signatures in <i>Gomorra</i>	41
Paola Dalla Torre, Paolo Fantozzi, Maurizio Naldi and Placido Pellegriti	
4. Women Behind the Scenes: A Data-Driven Approach to Analyse Female Labour in Italian Television	65
Valentina Re and Marica Spalletta	
5. Constructing an Open, Participatory Database on Gender (In)Equality in the Italian Film Industry: Methodological Challenges	87
Rosa Barotsi, Mariagrazia Fanchi and Matteo Tarantino	
6. Women in Polish TV Series: Qualitative and Quantitative Analysis	105
Andrzej Meler, Beata Królicka and Radosław Sojak	
7. Queer and Mainstream: Tracing <i>Heartstopper's</i> Reception via Digital Tools	125
Marta Boni	
8. TV Reception via Social Media Analysis: The Case of <i>Doc – Nelle tue mani</i>	141
Stefania Antonioni and Dom Holdaway	
Contributors	163

1. A Doubtful Handshake

The Promising Frontier of Data-Driven Film and Media Studies

Giorgio Avezzù and Marta Rocchi

Charles Marion Russell was a prolific American artist who, in the early 1900s and the first decades of the new century, created powerful paintings depicting the wild west of the United States. These paintings also had some influence on the cinematic imagery of the western genre. One of his beautiful watercolours from 1910, titled “A Doubtful Handshake”, shows a cowboy and a Native American leaning toward each other, each on their own horse, to shake hands, albeit with some suspicion on both sides, against the backdrop of a desolate frontier landscape and snow-capped mountains under an off-white sky. The artist later returned to the same subject, and a similar oil painting is part of the British Royal Collection under the title “Left Handshake Is the Safest” – the right hand is ready with a rifle, just in case.

It seems to us that the image of the doubtful handshake is an appropriate way to represent the encounter that we also hope to promote with this volume, as well as the difficulties that may arise, such as mistrust and resistance. We are referring to the intersection between film and media studies, on the one hand, and the culture of data and data-driven analysis tools, on the other. In other words, we are ultimately talking about a particular opportunity to bring together the “two cultures” (Snow 1959), given that studies on cinema and audiovisual media mostly come from the humanities tradition, while it is generally the STEM disciplinary fields that deal with data. We do not want to force the metaphor too much, and establish who are the cowboys and who are the Indians, and who keeps his right hand on the rifle, but certainly the encounter we are referring to represents a frontier theme, which has become essential today as a result of the innovations introduced by the digital revolution in our field of study, in the realm of the objects that we are used study.

Certainly, it is an issue that is impossible to ignore today, and one that we feel is urgent to confront, even more so now than in the past. However, it is not a theme that originated with the digital era: one could attempt a sort of archaeology of the approaches that have tried to base media studies on the most solid foundations of the so-called hard sciences. Among these, attempts to “datify” the textual aspects of films for their formal and stylistic analysis are well known. These are attempts that had already been anticipated by some previous experiments, but that began to gain some diffusion from the mid-1970s. Think of the analysis of average shot length to investigate the stylistic signatures of films directed by different directors, or to survey cutting rates in different periods of film history (e.g., the 1940s and 1950s) (Salt 1974, 1983, Tsivian 2009, Heftberger 2018, see also Bordwell 2006). Similar quantitative analyses, as well as those on the scale of shot and camera movement, have been carried out for decades, literally manually, without recourse to automated tools.

The strength of objectivity of the quantitative data of “cinematics”, its apparent self-evidence, has perhaps sometimes led to sacrificing the depth of interpretation of the data itself. This has also been observed recently by Franco Moretti – a scholar who cannot be suspected of being hostile to a “distant reading” approach – precisely in relation to certain perhaps hasty conclusions on alleged trends observable from data that would show the shortening of the duration of shots in American cinema over the decades (Moretti 2022, about Cutting and Candan 2015). It seems to us that even this kind of objection basically interprets the old opposition between the “two cultures” and the difficulties of communication between one and the other.

The contrast between more traditional approaches and data-driven analysis in cinema studies, on the other hand, is no doubt well understood by anyone who reads the pages of a paradigmatic scholar like Barry Salt (1983), particularly the first chapters of his *Film Style and Technology*, “the first and only history of motion picture style”. One cannot help but notice in those pages a certain arrogance on the part of the author, who boasts a PhD in theoretical physics and believes in the objectivity of what he calls “real sciences”, and an explicit aversion to the humanities in general and their alleged charlatanism. An aversion that was motivated, for that matter, also by the hostility reserved for Salt himself by many “traditional” film scholars as well as by all those major film journals that denied him, “with no reason given”, the publication of earlier versions of his studies. These same tensions seem to us to often continue to persist even today, but perhaps it

would be a case of finally tempering or dissolving them – at least this is what we would like to contribute to doing – by reconciling the scientific reliability of the method and the theoretical fecundity of the research question, and harmonizing the objectivity of analysis with the depth of interpretation. As Charles Percy Snow (1959) himself said, “This polarization is sheer loss to us all. To us as people, and to our society”.

This book focuses on the datafication of audiovisual media content and audiences, as well as data-driven methods and methodologies for studying films and TV series. While quantitative perspectives applied to narrative audiovisual texts and audience behavior have existed in the history of film and media studies, as noted above, they have often been met with resistance from proponents of more traditional approaches. However, there is an increasingly established tradition of quantitative studies at the international level. Influential researchers have emphasized the potential of these studies, as well as the gap between the possibilities offered by digital data and computational tools, and their actual application in the field of media studies (Noordegraaf 2016). In this context, various methodological approaches have been developed to uncover patterns, trends, and characteristics of audiovisual products. These approaches include processing textual objects through multimodal techniques, analyzing social discursivity using automated software, and evaluating production and consumption processes through social network analysis and modeling techniques. Each of these research projects is grounded in specific tools for collecting and analyzing audiovisual media data, using statistical and computational techniques. The aim of this volume is to draw attention to and facilitate discussion on systematic methods for conducting data-driven research in film and media studies. Specifically, it delves into the insights that data-driven approaches can offer for the study of contemporary audiovisual products. We share the belief that “a unification of the field might come about not in the first instance, through a standardization of the elementary research procedures, but in the final instance, through an open-ended application and comparison of multiple methodology” (Jensen 2020).

The volume features some of the contributions from the program of the thirteenth edition of the Media Mutations conference, titled *Audiovisual Data: Data-Driven Perspectives for Media Studies*, which was held in Bologna on October 6th and 7th, 2022. The conference, as well as this book, was intended as an opportunity both to investigate *what we can do with data* and the analytical and interpretative possibilities at stake, and to reflect on

what data can be (and what kind of data we must deal with), in order to problematize the possible limits of such approaches to the study of audiovisual narrative media. To gain a better understanding of what can be done with data, and the potential of data itself, we will briefly explore some of the most popular and promising areas of research in media studies. This will involve considering several interesting methods and approaches to data-driven research that are relevant to the contemporary landscape of media studies.

One of the key areas of inquiry related to data-driven research in media studies is media content analysis. This method has a rich history, which Macnamara (2018) details in his work. Initially, media content analysis was developed as a systematic method to study mass media, specifically propaganda (Lasswell 1927). Over time, content analysis became an important research method for studying portrayals of violence, racism, and representations of women in television programming and films. This is because such symbolic content has the potential to influence attitudes and behavior. These approaches are still prevalent today. For example, Greenberg and colleagues (2009) compared food and drink consumption across different television program genres using content analysis. Himes and Thompson (2017) analyzed the specific gender, age, and types of fat stigmatization frequently depicted in movies and television shows using content analysis. Chapoton and colleagues (2020) utilized content analysis to identify patterns and frequencies of occurrences linked to alcohol within TV series. Content analysis has also been implemented to examine depictions of gender (Sink and Mastro 2017, González-de-Garay et al. 2020) and immigration (Ramos et al. 2019) on primetime television and to investigate narrative features in US medical TV series (Rocchi and Pescatore 2022). By examining media content, researchers can gain valuable insights into how media might shape attitudes and behaviors related to important social issues.

Despite its long history, content analysis continues to be a valuable tool in media studies research, as it enables researchers to make replicable and valid inferences by interpreting and coding audiovisual data, and converting qualitative data into quantitative data. In this context, computer-assisted analysis has significantly improved the process of content analysis, allowing researchers to boost the conversion of qualitative to quantitative data. Moreover, with the advent of machine learning techniques, such as topic modeling and natural language processing, data-driven studies in several areas of interest for media studies have become possible. For instance, studies can recognize characters from scripts, dialogues, captioned images,

and video material (Cour et al. 2010, Bojanowski et al. 2013, Parkhi et al. 2015, Huang et al. 2020, Nagrani and Zisserman 2018), conduct linguistic analysis on TV series dialogue (Bednarek 2018, 2020, Tukachinsky et al. 2022), or create screenplay summaries (Gorinski et al. 2015, Papalampidi et al. 2020).

An interesting perspective of media content analysis is its intersection with the study of diversity issues. Indeed, by conducting qualitative and quantitative content analysis, it is possible to study on-screen diversity (Sink and Mastro 2017, Ramos et al. 2019, González-de-Garay et al. 2020) and relate it to one of the components of off-screen diversity, which is the diversity of the workforce within individual media outlets, also known as “source diversity” (Napoli 1999). Quantitative investigations have already explored who is responsible for on-screen representation within the production context. For instance, Loist and Prommer (2019) conducted data collection on different sources, such as funding and box office, and empirically highlighted how gender imbalances in the German film industry contribute to the unsustainability and precarity of women creative professionals’ careers, such as directors, producers, cinematographers, and scriptwriters. To monitor and assess large-scale off-screen diversity within the audiovisual production industry, one of the most useful tools is to use IMDb as a data source (Fanchi and Tarantino 2019, 2020, Verhoeven et al. 2019, 2020), although this approach has some limitations. Professional data, product information, and production details can also be useful for exploring broader phenomena that are usually difficult to investigate in their totality, using Social Network Analysis (SNA) techniques. For instance, Verhoeven and colleagues (2020) applied network science to the question of gender inequality within the Australian, German, and Swedish film industries, and showed that the expansion of collaborative sector networks between men and women is the most likely way to generate behavioral changes that improve network openness and address gender inequality. SNA techniques have also been useful in investigating the production of US TV series, between capital and labour strategies (Ruffino and Brembilla 2016) and cinema of “cultural interest” (Holdaway 2016). Furthermore, the interaction among audiovisual characters is a growing area of interest and can be analyzed in relation to textual analysis to predict the performance of TV series (Fronzetti Colladon and Naldi 2019, Hunter 2019).

Artificial intelligence (AI) is certainly a crucial area of investigation in data-driven research in media studies. Indeed, with the increasing signifi-

cance of data analytics and artificial intelligence strategies, many creative industries are now leveraging these technologies to improve their production processes (Anantrasirichai and Bull 2022). As a result, the number of data-driven studies in film, media, and sound studies has also grown substantially. Artificial intelligence is now being used for a range of tasks, including voice, text and image recognition (Fischer et al. 2011, Doukhan et al. 2018, Nagrani and Zisserman 2018), archives practices (Colavizza et al. 2021, De Quintana et al. 2021), decision-making support in the film industry (Ghiassi et al. 2015), and content-based predictive models of consumer behavior (Toubia et al. 2019, 2021). With the increasing use of media data, AI is being employed in various areas for the production and automatic generation of narrative audiovisual content, such as video editing, clip selection, filming, and screenplay writing. For instance, Smith and colleagues (2017) used AI to create a real horror movie trailer by modeling the key defining components that characterize horror movie trailers, while Bost and colleagues (2019) developed a tool for the automatic generation of video summaries of TV series that models the dynamics of the plot and extracts relevant sequences to create character-oriented video summaries. Derda (2022) has recently identified a new trend in the audiovisual production process, where artificial intelligence is now being used to support all phases of production. Traditionally, audiovisual production has been composed of three main phases: pre-production, production, and post-production. However, with AI-Supported Production Process, the incorporation of evaluations and feedback from the collection, analysis, and application of consumer data can now be integrated into all phases of audiovisual production, from initial conceptualization to final distribution. Derda (2022: 12) explains that this approach involves a “repetitive, or even continuous, evaluation process of the performance of the content and analysis of audience feedback”, resulting in a shift in the nature of the production process from a relatively linear to a cyclical one. Artificial intelligence is revolutionizing the creative industries by enabling producers to gain valuable insights into consumer behavior and preferences, which can inform their creative strategies and deliver content that resonates more effectively with audiences.

The intersection between data analytics and theory has been explicitly sought in research that addresses audiovisual media within the disciplinary fields of Business and Economics. These studies share the same objectives that marketing departments of major film studios and television networks (especially overseas) have pursued since the 1980s and 1990s, with a certain

degree of optimism. The underlying idea is that audience appreciation of audiovisual products is not random, and success is not merely a matter of chance. Therefore, by studying data, it is possible to identify patterns and regularities, and even build models that can support various phases of production and distribution (Hennig-Thurau and Houston 2019). It would not be true, as is often argued in every sector of the entertainment industry, that “nobody knows anything” about what it takes for a certain product to become a hit, and that it is only right and inevitable that every strategic decision in this field is made instinctively, “by gut feeling”, following intuition. In fact, a quantitative study of audiovisual texts on the one hand, and of markets and consumption on the other, would make it possible to identify factors that attract viewers. This, in turn, can “substantially [increase] the probability of success in the entertainment industry” (Hennig-Thurau and Houston 2019) or at least predict the reception that certain content will have based on how it was made, promoted, and distributed.

If the proponents of *Entertainment Science*, as it is called, are mainly business scholars and entertainment economists, it is because it was often the companies themselves that were particularly familiar with data. And they still are, usually much more so than film and media studies academics, at least. The film industry typically studies data on box office revenue and attendance, and on the number of theatres in which films are distributed, also in relation to geographical and temporal dimensions, and integrates this data with interviews and qualitative research to better understand the characteristics of those who watch films in theatres. The television industry is even more familiar with data and has at its disposal even more granular data derived from user panels employed by consumer research and market measurement companies, which provide very precise details on the demographics of viewers, such as their socio-economic class. Digitalization has led, in Italy as elsewhere, to the integration of the classic “panel based” paradigm in the measurement of television audience data with a “census” paradigm. Total television viewing is no longer solely inferred from the behaviors of the representative panel considered by the measurement company and then projected onto the entire population through statistical data expansion factors, but is also described through tracking all individual devices enabled for viewing via IP protocol that consume content through apps or browsers being surveyed, such as tablets, smartphones, smart TVs, computers, etc., both live and on-demand, at home or away. This has introduced new metrics, and a whole series of challenges to reconcile “traditional” data with that

of the new “everytime, everywhere, everydevice” digital TV (Guarnaccia 2022 offers some insights on this topic).

From much of these challenges and data, however, academics are usually excluded, with rare exceptions (e.g., Scaglioni 2021, 2022, Thurman 2021, see also Avezzù 2022 on the geographies of film and TV consumption). This exclusion hinders the communication that we would like to see, which is not only between the humanities and the hard sciences, but also that which should result from the encounter between the university and the audiovisual industry. Notably, the use of data and of tools for their (geo)visualization, for example to describe cinemagoing behaviors, seems to be more frequent – within film studies – in reference to distant decades of film history than in reference to the contemporary scenario. This is the case, among others, with certain contributions of *New Cinema History* and other research on national audience histories (Maltby, Walker and Walsh 2014, Treveri Gennari et al. 2020, Baptist, Noordegraaf and van Oort, 2021), or even with attempts to elaborate new tools for measuring film popularity (Sedgwick 2009). Without detracting from all these contributions and the various strands of research they represent, which we believe to be fundamental, it is inevitable to note a greater propensity to experiment with new tools and methodologies on objects of study from which a certain historical distance separates us, while applications on more contemporary objects seem rarer, not least because of a greater scarcity of data, which may seem to be paradoxical.

One field where both the relevance of data and the lack of its availability for academic research is particularly evident is that of video-on-demand (VOD) platforms. This is especially true for subscription-based platforms, which are excluded from any official survey perimeter by measurement companies in Italy and elsewhere. As a result, they do not employ any certified or publicly shared data or metrics. This is still a very loosely regulated area, where individual players, such as Netflix, are currently free to construct their own metrics and use them at will, primarily for promotional purposes, and without any possibility of third-party verification or control. Apart from their promotional and sometimes ambiguous use, video-on-demand platforms use data for content intelligence, that is, to decide what content to produce, and to feed their recommendation systems, and thus to profile content offerings according to users’ tastes. Subscription-based VOD operators have often extolled such recommender systems as fundamental to their own services (Gomez-Uribe and Hunt 2015, Hallinan and Striphos 2016), and this has solicited some interest from the academic community,

but these are, of course, objects of study that are out of reach for academics, and can only be approached as black boxes. Various attempts by media scholars to “expose” or “reverse engineer” Netflix’s recommendation systems – or rather, its “algorithm” – have more than anything else demonstrated the difficulty of studying something from which one is excluded, and interrogating data and data-driven logics to which one is not granted access (Pajkovic 2021, Gaw 2022).

As we have seen through this swift and non-exhaustive survey, the application of data-driven methods to the study of audiovisual products opens up a vast array of perspectives. It also requires careful reflection on the nature of data itself: what do we talk about when we talk about data? It is also perhaps in this area that scholars from the humanities can offer an important contribution by questioning and problematizing the data. What can we do with the data today, and what can we study better than before? What research can add something new to the classic media studies debates, and in what specific subfields? What new perspectives can we gain access to that were previously unavailable to us? What data and tools do we need to achieve our goals? On the other hand, what are the limitations of data-driven methods, and are there any unnecessary or misleading approaches that cannot be brought into dialogue with traditional strands of research in audiovisual studies? Furthermore, what are the media industries already doing with data, and can we learn anything from them? As academic scholars can we contribute anything useful to them?

Below, we provide a brief overview of the contributions collected in this volume.

Mirko Degli Esposti and Guglielmo Pescatore, who collaborated with us on the curation of the 13th edition of the Media Mutations conference, authored the first article in this collection. Their essay delves into the use of data-driven approaches and computational techniques in television studies, while also addressing the challenges posed by the complex open systems of TV series that interact with external factors. The article further examines the possibilities of generative AI and the use of deep multimodal algorithms to create new content.

The second contribution of this collection focuses on the automatic content analysis of dialogues of the TV series *Gomorra*. Paola Dalla Torre and colleagues utilize a data-driven approach to uncover the correlation between

a character and their lexical signature. They employ machine learning algorithms to process the show's standardized transcripts, analyzing various features, including word presence, lexicon similarity and variety, character presence, and character identification.

In the third contribution, Valentina Re and Marica Spalletta investigate women's employment in key behind-the-scenes roles in Italian TV crime dramas broadcast from Fall 2015 to Summer 2022. The data reveals that there are stronger gender inequalities in the Italian television production industry than in the European averages for TV fiction and the Italian film sector, despite some positive trends.

Barotsi and colleagues examine the relationship between the “ecology of knowledge”, data, and gender equality policies, using the CENTRIC project as a starting point. They discuss the methodological challenges in creating an open and participatory research platform on gender inequality in the audiovisual industry, which can serve as a powerful tool for promoting inclusiveness and cultural pluralism.

Andrzej Meler and colleagues analyze how modern Polish TV series attempt to challenge and reinforce stereotypes about women through their depiction of female characters. They employ a multi-layered coding approach to measure the contexts in which female characters are presented and a qualitative analysis to identify the stereotypes or anti-stereotypes presented in the series.

Marta Boni's essay delves into the potential of open-access digital tools for studying media content as a space of discourse. By integrating both quantitative and qualitative analysis, Boni argues that such an approach can reveal interesting features that may not be identifiable through qualitative analysis alone and the adaptation and testing of tactics for the specific object of study can provide deeper understanding.

Stefania Antonioni and Dom Holdaway employ data-driven methods to analyze the reception practices and discussions surrounding the Italian medical drama series *Doc – Nelle tue mani* on social media platforms. Their study examines user comments, discussions, and visual content on Instagram and YouTube, utilizing both quantitative and qualitative analyses of scraped data. In their analysis, the authors explore the limitations and challenges of data-driven approaches in such studies.

BIBLIOGRAPHY

- Anantrasirichai, Nantheera and David Bull (2022). "Artificial Intelligence in the Creative Industries: A Review." *Artificial Intelligence Review* 55: 589-656.
- Avezzi, Giorgio (2022). *L'Italia che guarda. Geografie del consumo audiovisivo*. Roma: Carocci.
- Baptist, Vincent, Julia Noordegraaf and Thunnis van Oort (2021). "A Digital Toolkit to Detect Cinema Audiences of the Silent Era: Scalable Perspectives on Film Exhibition and Consumption in Amsterdam Neighbourhoods (1907-1928)." *Studies in European Cinema* 18(3): 252-73. <https://doi.org/10.1080/17411548.2021.1930974>.
- Bednarek, Monika (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge: Cambridge University Press.
- Bednarek, Monika (2020). "On the Usefulness of the Sydney Corpus of Television Dialogue as a Reference Point for Corpus Stylistic Analyses of TV Series." In *Telecinematic Stylistics*, edited by Christian Hoffman and Monika Kirner-Ludwig, 39-62. London: Bloomsbury.
- Bojanowski, Piotr, Francis Bach, Ivan Laptev, Jean Ponce, Cordelia Schmid and Josef Sivic (2013). "Finding Actors and Actions in Movies." In *Proceedings of the IEEE international conference on computer vision*, 2280-87.
- Bordwell, David (2006). *The Way Hollywood Tells It: Story and Style in Modern Movies*. Berkeley: University of California Press.
- Bost, Xavier, Serigne Gueye, Vincent Labatut, Martha Larson, Georges Linares, Damien Malinas and Raphaël Roth (2019). "Remembering Winter was Coming: Character-Oriented Video Summaries of TV Series." *Multimedia Tools and Applications* 78: 35373-99. <https://doi.org/10.1007/s11042-019-07969-4>.
- Chapoton, Boris, Anne-Laure Werlen and Véronique Regnier Denois (2020). "Alcohol in TV Series Popular with Teens: A Content Analysis of TV Series in France 22 Years after a Restrictive Law." *European Journal of Public Health* 30(2): 363-68. <https://doi.org/10.1093/eurpub/ckz163>.
- Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens and Julia Noordegraaf (2021). "Archives and AI: An Overview of Current Debates and Future Perspectives." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15(1): 1-15. <https://doi.org/10.1145/3479010>.
- Cour, Timothee, Benjamin Sapp, Akash Nagle and Ben Taskar (2010). "Talking Pictures: Temporal Grouping and Dialog-Supervised Person Recognition." In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1014-21.

- Cutting, James E. and Ayse Candan (2015). “Shot Durations, Shot Classes, and the Increased Pace of Popular Movies.” *Projections* 9(2): 40-62. <https://doi.org/10.3167/proj.2015.090204>.
- De Quintana, Eugenio López and Antonio León Carpio (2021). “Artificial Intelligence for a Role Change in Television Archives: The Atresmedia–Etiqmedia Experience.” *Journal of Digital Media Management* 10(2): 177-87.
- Derda, Izabela (2022). “‘Did you Know that David Beckham Speaks Nine Languages?’: AI-Supported Production Process for Enhanced Personalization of Audio-Visual Content.” *Creative Industries Journal* 1-16. <https://doi.org/10.1080/17510694.2021.2025001>.
- Doukhan, David, Géraldine Poels, Zohra Rezgui and Jean Carrive (2018). “Describing Gender Equality in French Audiovisual Streams with a Deep Learning Approach.” *VIEW Journal of European Television History and Culture* 7(14): 103-22. <https://doi.org/10.25969/mediarep/14757>.
- Fanchi, Mariagrazia and Matteo Tarantino (2019). “Donne e reti creative nel cinema italiano (2004-2016).” *Economia della Cultura* 29(4): 523-30. <https://doi.org/10.1446/96648>.
- Fanchi, Mariagrazia and Matteo Tarantino (2020). “Has Digital Transformation Impacted Gender Imbalance in Italian Cinema? A Data Analysis of Creative Clusters 2004–2016”. In *Digital Transformation in the Cultural and Creative Industries*, edited by Marta Massi, Marilena Vecco and Yi Lin, 162-184. London: Routledge.
- Fischer, Mika, Hazım Kemal Ekenel and Rainer Stiefelwagen (2011). “Person Re-Identification in TV Series Using Robust Face Recognition and User Feedback.” *Multimedia Tools and Applications* 55: 83-104. <https://doi.org/10.1007/s11042-010-0603-2>.
- Fronzetti Colladon, Andrea and Maurizio Naldi (2019). “Predicting the Performance of TV Series through Textual and Network Analysis: The Case of *Big Bang Theory*.” *PLoS ONE* 14(11): e0225306. <https://doi.org/10.1371/journal.pone.0225306>.
- Gaw, Fatima (2022). “Algorithmic Logics and the Construction of Cultural Taste of the Netflix Recommender System.” *Media, Culture & Society* 44(4): 706-25. <https://doi.org/10.1177/01634437211053767>.
- Ghiassi, Manoochehr, David Lio and Brian Moon (2015). “Pre-Production Forecasting of Movie Revenues with a Dynamic Artificial Neural Network.” *Expert Systems with Applications* 42(6): 3176-93. <https://doi.org/10.1016/j.eswa.2014.11.022>.
- Gomez-Urbe, Carlos A. and Neil Hunt (2015). “The Netflix Recommender System: Algorithms, Business Value, and Innovation.” *ACM Transactions on Management Information Systems* 6(4) art. 13: 1-19. <https://doi.org/10.1145/2843948>.
- González-de-Garay, Beatriz, María Marcos-Ramos and Carla Portillo-Delgado (2020). “Gender Representation in Spanish Prime-Time TV Series.” *Feminist Media Studies* 20(3): 414-33. <https://doi.org/10.1080/14680777.2019.1593875>.
- Gorinski, Philip and Mirella Lapata (2015). “Movie Script Summarization as Graph-Based Scene Extraction.” In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1066-76.
- Greenberg, Bradley S., Rosaen Sarah F., Worrell Tracy R., Salmon Charles T. and Julie E. Volkman (2009). “A Portrait of Food and Drink in Commercial TV Series.” *Health Communication* 24(4): 295-303. <https://doi.org/10.1080/10410230902889233>.

- Guarnaccia, Fabio (ed.) (2022). *Metrix. Viaggio all'ultima frontiera delle metriche. Link. Idee per la televisione* 28.
- Hallinan, Blake and Ted Striphas (2016). "Recommended for You: The Netflix Prize and the Production of Algorithmic Culture." *New Media & Society* 18(1): 117-37.
- Heftberger, Adelheid (2018). *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work*. Cham: Springer.
- Hennig-Thurau, Thorsten and Mark B. Houston (2019). *Entertainment Science: Data Analytics and Practical Theory for Movies, Games, Books, and Music*. Cham: Springer.
- Himes, Susan M. and J. Kevin Thompson (2007). "Fat Stigmatization in Television Shows and Movies: A Content Analysis." *Obesity* 15(3): 712-18. <https://doi.org/10.1038/oby.2007.635>.
- Holdaway, Dominic (2016). "La rete sociale del cinema di interesse culturale." In *Il cinema di Stato. Finanziamento pubblico ed economia simbolica nel cinema italiano contemporaneo*, edited by Marco Cucco and Giacomo Manzoli, 127-69. Bologna: Il Mulino.
- Huang, Qingqiu, Lei Yang, Huaiyi Huang, Tong Wu and Dahua Lin (2020). "Caption-Supervised Face Recognition: Training a State-of-the-Art Face Model without Manual Annotation." In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII*, 139-55. Cham: Springer International Publishing.
- Hunter, Starling (2019). "Predicting Nielsen Ratings from Pilot Episodes Scripts: A Content Analytical Approach." *SERIES – International Journal of TV Serial Narratives* 5(1): 9-21. <https://doi.org/10.6092/issn.2421-454X/9079>.
- Jensen, Klaus Bruhn (2020). *A Handbook of Media and Communication Research: Qualitative and Quantitative Methodologies*. London: Routledge.
- Lasswell, Harold Dwight (1927). *Propaganda Techniques in the World War*. New York: Knopf.
- Loist, Skadi and Elizabeth Prommer (2019). "Gendered Production Culture in the German Film Industry." *Media Industries* 6(1): 95-115.
- Macnamara, Jim (2018). "Content Analysis." In *Mediated Communication*, edited by Philip M. Napoli, 191-212. Berlin-Boston: De Gruyter. <https://doi.org/10.1515/9783110481129-012>.
- Maltby, Richard, Dylan Walker and Mike Walsh (2014). "Digital Methods in New Cinema History." In *Advancing Digital Humanities*, edited by Katherine Bode and Paul Longley Arthur, 95-112. London: Palgrave Macmillan.
- Mbunge, Elliot, Stephen Gbenga Fashoto and Happyson Bimha (2022). "Prediction of Box-Office Success: A Review of Trends and Machine Learning Computational Models." *International Journal of Business Intelligence and Data Mining* 20(2): 192-207. <https://doi.org/10.1504/IJBIDM.2022.120825>.
- Moretti, Franco (2022). *Falso movimento. La svolta quantitativa nello studio della letteratura*. Milano: Nottetempo.
- Nagrani, Arsha and Andrew Zisserman (2018). "From Benedict Cumberbatch to Sherlock Holmes: Character Identification in TV Series without a Script." *arXiv preprint arXiv:1801.10442*. <https://doi.org/10.48550/arXiv.1801.10442>.
- Napoli, Philip M. (1999). "Deconstructing the Diversity Principle." *Journal of Communication* 49(4): 7-34. <https://doi.org/10.1111/j.1460-2466.1999.tb02815.x>.

- Noordegraaf, Julia (2016). “Computational Research in Media Studies: Methodological Implications.” *Kwalon* 21(1): 52-59. <https://doi.org/10.5117/2016.021.001.007>.
- Pajkovic, Niko (2021). “Algorithms and Taste-Making: Exposing the Netflix Recommender System’s Operational Logics.” *Convergence: The International Journal of Research into New Media Technologies* 28(1): 214-235. <https://doi.org/10.1177/13548565211014464>.
- Papalampidi, Pinelopi, Frank Keller, Lea Frermann and Mirella Lapata (2020). “Screenplay Summarization Using Latent Narrative Structure.” *arXiv preprint arXiv:2004.12727*. <https://doi.org/10.48550/arXiv.2004.12727>.
- Parkhi, Omkar M., Esa Rahtu and Andrew Zisserman (2015). “It’s in the Bag: Stronger Supervision for Automated Face Labelling.” *International Conference on Computer Vision*.
- Ramos, María Marcos, Beatriz González de Garay and Carla Portillo Delgado (2019). “The Representation of Immigration in Contemporary Spanish Primetime TV Series.” *Revista Latina de Comunicación Social* 74: 285-307.
- Rocchi, Marta and Guglielmo Pescatore (2022). “Modeling Narrative Features in TV Series: Coding and Clustering Analysis.” *Humanities and Social Sciences Communications* 9, 333. <https://doi.org/10.1057/s41599-022-01352-9>.
- Ruffino, Marco and Paola Brembilla (2016). “Narrative Ecosystems through the Network Analysis Lens. Step One: The Production of US TV Series, between Capital and Labor Strategies.” *SERIES –International Journal of TV Serial Narratives* 2(1): 55-68. <https://doi.org/10.6092/issn.2421-454X/6164>.
- Salt, Barry (1974). “Statistical Style Analysis of Motion Pictures.” *Film Quarterly* 28(1): 13-22. <https://doi.org/10.2307/1211438>.
- Salt, Barry (1983). *Film Style and Technology: History and Analysis*. London: Starword.
- Scaglioni, Massimo (2021). *La televisione nella pandemia. Intrattenimento, fiction, informazione e sport nell'anno del Covid-19. Annuario 2021*. Roma: Carocci.
- Scaglioni, Massimo (ed.) (2022). *Total TV. Intrattenimento, fiction, informazione e sport fra broadcasting e streaming. Annuario 2022*. Milano: Certa.
- Sedgwick, John (2009). “Measuring Film Popularity: Principles and Application.” In *Digital Tools in Media Studies: Analysis and Research: An Overview*, edited by Michael Ross, Manfred Grauer, Bernd Freisleben, 43-54. Bielefeld: Transcript Verlag.
- Sink, Alexander and Dana Mastro (2017). “Depictions of Gender on Primetime Television: A Quantitative Content Analysis.” *Mass Communication and Society* 20(1): 3-22. <https://doi.org/10.1080/15205436.2016.1212243>.
- Smith, John R., Dhiraj Joshi, Benoit Huet, Winston Hsu and Jozef Cota (2017). “Harnessing AI for Augmenting Creativity: Application to Movie Trailer Creation.” In *Proceedings of the 25th ACM International Conference on Multimedia*, 1799-1808.
- Snow, Charles P. (1959). *The Two Cultures and the Scientific Revolution*. London: Cambridge University Press.
- Thurman, Neil (2021). “When a TV Channel Reinvents itself Online: Post-Broadcast Consumption and Content Change at BBC Three.” *Convergence: The International Journal of Research into New Media Technologies* 27(2): 291-312. <https://doi.org/10.1177/1354856520967773>.
- Toubia, Olivier, Garud Iyengar, Renée Bunnell and Alain Lemaire (2019). “Extracting

- Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption.” *Journal of Marketing Research* 56(1): 18-36.
- Toubia, Olivier, Jonah Berger and Jehoshua Eliashberg (2021). “How Quantifying the Shape of Stories Predicts their Success.” *Proceedings of the National Academy of Sciences* 118(26): e2011695118. <https://doi.org/10.1073/pnas.2011695118>.
- Traverri Gennari, Daniela, Catherine O’Rawe, Danielle Hipkins, Silvia Dibeltulo and Sarah Culhane (2020). *Italian Cinema Audiences: Histories and Memories of Cinemagoing in Post-War Italy*. New York: Bloomsbury.
- Tsivian, Yuri (2009). “Cinematics, Part of the Humanities’ Cyberinfrastructure.” In *Digital Tools in Media Studies: Analysis and Research: An Overview*, edited by Michael Ross, Manfred Grauer and Bernd Freisleben, 93-100. Bielefeld: Transcript Verlag.
- Tukachinsky, Forster, Rebecca Riva, Dana Mastro and Marko Dragojevic (2022). “Linguistic Characteristics of Interracial Interactions on Primetime TV: A Quantitative Content Analysis.” *Western Journal of Communication* (2022): 1-25. <https://doi.org/10.1080/10570314.2022.2146462>.
- Verhoeven, Deb, Bronwyn Coate and Vejune Zemaityte (2019). “Re-Distributing Gender in the Global Film Industry: Beyond #MeToo and #MeThree.” *Media Industries Journal* 6(1): 135-55. <https://doi.org/10.3998/mij.15031809.0006.108>.
- Verhoeven, Deb, Katarzyna Musial, Stuart Palmer, Sarah Taylor, Shaukat Abidi, Vejune Zemaityte and Lachlan Simpson (2020). “Controlling for Openness in the Male-Dominated Collaborative Networks of the Global Film Industry.” *PloS ONE* 15(6): e0234460. <https://doi.org/10.1371/journal.pone.0234460>.

A DOUBTFUL HANDSHAKE: THE PROMISING
FRONTIER OF DATA-DRIVEN FILM AND MEDIA STUDIES



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.fa016619>.

In Giorgio Avezù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

2. Exploring TV Seriality and Television Studies through Data-Driven Approaches

Mirko Degli Esposti and Guglielmo Pescatore

◀ ABSTRACT

The article discusses the use of data-driven approaches in television studies, which has become possible due to the increasing availability of digital data. Computational techniques can be used to analyze cultural artifacts, gain insights into audience reactions to specific shows or episodes, and investigate patterns and trends in television programming over time. The article also highlights the challenges of analyzing television series, which are complex open systems that interact with external factors such as the production process, audience feedback, and cultural and social context. Content analysis, which involves qualitative and quantitative methods based on the coding process and data collection, can be used to analyze various elements of a TV series.

Generative AI is also discussed, which refers to the use of deep multimodal algorithms to generate new content such as images, speech, and text. Generative methods like Generative Adversarial Networks (GANs) and Stable Diffusion can create new content that is almost indistinguishable from real data. While generating videos is more challenging, Recurrent Neural Networks (RNNs) like LSTMs can capture the temporal dynamics of the scenes to create interesting and promising applications for complex, but short duration videos.

KEYWORDS

Television series; television studies; data-driven audiovisual analysis; generative AI; content analysis.

Datafication and Data-Driven Approaches in the Humanities

Data-driven approaches have become increasingly popular in the humanities for a variety of reasons. The growing availability of digital data means that researchers can analyze, model and interpret cultural artifacts like books, paintings, photographs, and music using computational tools and techniques. This has opened up new avenues for research and analysis that were not previously possible. One example is digital humanities, an interdisciplinary field that emerged from the use of computational methods to study history, literature, arts, and other forms of human culture.

Data-driven approaches allow for the processing of large amounts of data quickly and efficiently, which would be difficult or impossible to do manually. This makes it possible to identify patterns and trends in the data, leading to new insights and discoveries. Data-driven approaches are inherently interdisciplinary, drawing on techniques and methods from computer science, statistics, mathematical physics, and other fields. This has fostered new collaborations and partnerships between humanities scholars and researchers in other disciplines, leading to new perspectives and new areas of inquiry that were previously impossible or impractical to investigate. For example, it is now possible to use computational tools to analyze the spread of ideas and cultural trends across different geographic regions, using data drawn from social networks or other digital sources such as the Web and search engines. Data-driven approaches also allow research findings to be communicated in innovative ways. Interactive visualizations, digital archives, and multimedia presentations enable broader public engagement with humanities research and improve its dissemination.

Applications in Television Studies

The field of data-driven approaches has extended in recent years to the realm of television studies, particularly in the analysis of audience reception and engagement in television programming. By analyzing data from social media platforms like Facebook, Instagram, and especially Twitter researchers can identify trends in how audiences respond to specific shows or episodes. This approach can provide insights into how audiences interpret and react to various elements of a programme, such as characters, storylines, or themes (Antelmi et al. 2018, Burgess and Bruns 2015, Crespo-Pereira and Juanatey-Boga 2017, Ibrahim et al. 2018, Molteni and Ponce de Leon 2016). Social media has become a valuable source of data for analyzing audience reactions to television shows. There are several methods of using social media data to gain insights into audience responses. Sentiment analysis is one such approach, which involves analyzing the perception of social media posts related to a particular television show. By identifying patterns in audience reactions, such as which characters or storylines are most popular or controversial, sentiment analysis can provide valuable insights into how audiences interpret and respond to different aspects of a show (Antonakaki et al. 2021, Giachanou and Crestani 2016, Rocchi 2022a, 2022b, Scharl et al. 2016). Of particular interest is the use of social media as a second screen: second screen engagement allows viewers to share their opinions, reactions, and commentary with a broader audience and to engage in online live conversations around TV shows (Buschow et al. 2014, Pehlivan 2021, Williams and Gonlin 2017). Another approach is to analyze the use of hashtags related to a television show (Pilař et al. 2021). Both second screen and hashtag analysis can help to identify trends and themes in audience reactions by providing insights into how audiences are discussing the show on social media.

Network analysis involves mapping the relationships between different social media users and their interactions with each other. By identifying key influencers and opinion leaders within the audience, network analysis can provide insights into how opinions and reactions are shared and spread across social media (Hecking et al. 2017, Rocchi 2022b). Content analysis is a further method of analyzing social media data related to a television show. This involves analyzing the text and images of social media posts to gain insights into the specific themes and topics that audiences are discussing (Hagy et al. 2019). Text mining and Natural Language Processing (NLP)

can also be used to analyze large amounts of textual data, such as reviews, to identify patterns in audience responses. By identifying common themes and issues that emerge across different reviews, researchers can gain a deeper understanding of how audiences interpret and respond to cultural artifacts. A very useful tool for processing the data provided by content analysis is topic modeling, a linguistic analysis technique discussed below.

Another way that data-driven approaches are being applied in television studies is through the use of digital archives. Public archives like the Internet Archive's TV News (<http://archive.org/details/tv>) provide access to an enormous amount of historical television content that can be analyzed using computational techniques. This allows researchers to investigate patterns and trends in television programming over time, such as changes in genres, formats, or styles. This kind of approach has already enabled, in the field of film, the valorisation of archives and audiovisual materials from the perspective of enhancing cultural heritage and public history. Many projects have been developed, both internationally, such as i-media-cities (<https://www.imediacities.eu/>, Sala and Bruzzo 2019) and pilot projects dedicated to a single archive, such as PH-Remix (<http://www.labcd.unipi.it/ph-remix/>). Such projects can be developed to enhance the enormous amount of archival material available in the field of television.

Beyond analyzing the reception and history of television programming, data-driven approaches are also being employed to study the production and distribution of television shows. For instance, network analysis can be used to map the relationships between different producers, networks, and platforms, shedding light on the complex networks that underlie the production and distribution of television content (Ruffino and Brembilla 2016, Fanchi and Tarantino 2020).

Artificial Intelligence

Undoubtedly, the application of Artificial Intelligence (AI) has brought a new dimension to the analysis of audiovisual text, which involves scrutinizing textual information within audio and video content. AI systems have been developed with a range of techniques to extract and analyze text from audio and video content. Speech recognition involves using AI algorithms to transcribe spoken language into text, while image recognition utilizes AI algorithms to identify and classify objects and text within images. Additionally, NLP is employed to analyze and understand human language

in text that is embedded in audio and video content, such as transcripts, subtitles, and closed captions. The most common applications of AI in the field of audio-visuals involve automatic transcription and translation. AI systems can automatically generate closed captions for videos by analyzing the spoken content and transcribing it into text and translate subtitles in videos from one language to another. Speech-to-text transcription is possible through AI systems, enabling users to search and analyze spoken content from audio and video recordings. AI systems can analyze video content to identify key topics, themes, and sentiment expressed in the video. Artificial intelligence is often categorized into different types based on their functions and capabilities, like classification and prediction, but also generating new artificial content, as we will discuss.

Classification is a type of AI that involves assigning a label or category to a given input or data point. This task is accomplished by training the AI system to recognize patterns in the input data and using those patterns to make predictions about which category the input belongs to. For instance, an AI system can identify and classify images into different categories such as animals, vehicles, or objects. Another example of classification is sentiment analysis where an AI system can analyze text data and classify it as positive, negative, or neutral based on the sentiment expressed in the text.

Topic modeling is a powerful technique used in NLP and machine learning to uncover the underlying themes and topics within a corpus of text data. One of the most widely used algorithms for topic modeling is Latent Dirichlet Allocation (LDA) (Blei et al. 2003). LDA is based on the assumption that each document in a corpus contains a mixture of topics, and each word in a document is generated by one of those topics. This probabilistic model uses Bayesian inference to estimate the topic distributions, making it a powerful tool for discovering the topics that exist in a text corpus. Another popular algorithm used for topic modeling is Non-negative Matrix Factorization (NMF) (Lee and Seung 1999), which decomposes a term-document matrix into two lower-rank matrices representing the topics and their corresponding weights. Unlike LDA, NMF is a non-probabilistic algorithm that requires an unsupervised learning algorithm to reduce the dimensionality of data into lower-dimensional spaces. Other algorithms that are quite widely used and often derived from the previous two are the Hierarchical Dirichlet Process (HDP), the Correlated Topic Model (CTM), and the Structural Topic Model (STM), that allow for the modeling of topics as a function of covariates, such as time, authorship, or genre.

Recently, there has been a growing interest in the use of deep learning models for topic modeling, particularly with the emergence of powerful pre-trained language models such as BERT and GPT-3. These models have been shown to perform well in a variety of NLP tasks, including topic modeling. One example of a deep learning-based topic modeling algorithm is Bertopic (Grootendorst 2022), which uses BERT as a feature extractor and applies clustering techniques to group similar documents into topics. Bertopic has been shown to outperform traditional topic modeling algorithms such as LDA and NMF on several benchmark datasets (Egger and Yu 2022). While deep learning-based topic modeling algorithms have shown promising results, they also require large amounts of training data and computational resources. Furthermore, their black-box nature can make it difficult to interpret the resulting topics, which may limit their usefulness in certain applications. Nonetheless, the use of deep learning models for topic modeling is an active area of research with potential for further advancements in the field.

Prediction is an important AI technique in the audiovisual field that involves using historical data to forecast future events or outcomes. AI models are trained on past audiovisual data and can learn to recognize patterns and relationships in that data. The system then utilizes these patterns to make predictions about future audiovisual events. For instance, an AI model can analyze historical data on user behavior and predict which streaming platform a user is likely to use based on their viewing history. This information can help streaming platforms optimize their content recommendations and improve user engagement. Another example is content prediction, where AI models can analyze historical data on viewing patterns to predict which type of content is likely to be popular in the future. These predictive algorithms can help streaming platforms reduce churn rates, improve retention, and increase revenue.

These models use a variety of statistical techniques to analyze large datasets and identify relationships and patterns that can be used to predict future outcomes. In machine learning, predictive models are typically trained using supervised learning techniques, which involve providing the algorithm with labelled training data. The algorithm then uses this data to identify patterns and relationships between the input features and the output variable, and then generates a predictive model that can be used to make predictions on new, unseen data. Deep learning models, on the other hand, are a type of artificial neural network that are designed to learn and make predictions, partially inspired by how the human brain work. Deep

learning models typically use multiple layers of artificial neurons to analyze and interpret data, and are particularly effective at processing large, complex datasets.

The most important predictive application in the audiovisual domain is that of recommendation engines (Avezzi 2017, Schrage 2020). The algorithms used by recommendation engines for streaming platforms are typically complex and rely on various machine learning and/or deep learning techniques. The most commonly used approaches include collaborative filtering, content-based filtering and hybrid filtering. Collaborative filtering is based on the idea that users who have similar consumption habits are likely to have similar preferences for songs, movies or TV shows. Collaborative filtering looks at a user's past behavior, such as the movies they have watched and rated, and then recommends new content based on the viewing history of other users who have similar tastes. Content-based filtering uses features of the movies or TV shows themselves, such as genre, director, or actors, to recommend new content to users. Hybrid filtering combines collaborative filtering and content-based filtering to provide more accurate recommendations. These approaches can be used by implementing various machine learning algorithms or to extract the features of a deep learning model that identifies patterns and relationships between movies or TV shows and user behavior.

Data-Driven TV Series Analysis

TV series can be complex in a number of ways, and these complexities can pose both creative and analytical challenges. In a more formal way we might say that TV series are (multimodal, multiscale) open complex systems. One aspect of TV series' complexity is their narrative structure. TV series often feature multiple plotlines (Pescatore and Rocchi 2019, Rocchi and Pescatore 2022) that intersect and evolve over the course of many episodes, and on different time scales. This can create a sense of continuity and depth that is difficult to achieve in other storytelling formats, like movies. Several approaches have recently been introduced to use quantitative and automatic data extraction and analysis methods for understanding and revealing these complexities in TV series and movies, sometimes using the temporal and statistical properties of shoots and scenes (Cutting et al. 2010, Guha et al. 2015, Huang et al. 2018, Miech et al. 2019, Rao et al. 2020). This also poses a challenge for the so-called *automated summarization*, an-

other interesting problem in TV series' quantitative analysis (Chen et al. 2004, Bost et al. 2019).

Another aspect of TV series' complexity is in their character development. TV series often feature large ensemble casts with complex relationships and motivations (Pescatore et al. 2014, Pescatore and Innocenti 2018). Characters may evolve over time, revealing new facets of their personality and backstory as the series progresses. A very instructive and interesting approach to quantitative *character identification*, without the script, is found in Nagrani and Zisserman (2018), but other approaches also appear in Zhang and colleagues (2008), Maldi and colleagues (2011), Tapaswi and colleagues (2012), and Naldi and Dalla Torre (2022). Restricted to audio, this becomes the well-known problem of *diarization*, where algorithms are developed to learn how to predict who is talking at a given time (Bredin and Gelly 2016, Skowron et al. 2016, Park et al. 2021).

TV series can also be complex in terms of their thematic content. An interesting and relatively new approach to the analysis of television series is that of content analysis, a mix of qualitative and quantitative methods based on the coding process and data collection. Content analysis can be used to quantify and analyze different elements of a TV series, such as formulaic aspects of a particular genre, e.g., legal drama (Rocchi and Sonogo 2022), the frequency of certain types of scenes or dialogue, the emotions portrayed by characters, or the cultural representations depicted in the show. It can also be used to identify patterns and trends over time, such as changes in the depiction of certain social issues or the portrayal of certain types of characters (Rocchi and Anselmeti 2021). Many series tackle complex social and political issues and explore nuanced moral and ethical questions. TV series are not only complex in terms of their internal narrative structures and character development, but they are also open systems that interact with a range of external factors (Pescatore 2018), such as the production process, the audience, and the broader cultural and social context. One way in which TV series interact with the production process is through the development and execution of the series itself. TV series are typically produced by a team of writers, directors, producers, and other professionals, who work together try to create a coherent and engaging story arc. The production process can have a significant impact on the final product, as decisions around casting, editing, and marketing can all influence the way the series is received by audiences. TV series are designed to be consumed over a period of time, and as such, they create a relationship between the audience and the story.

Viewers become invested in the characters and their fates and may engage in a range of activities to deepen their understanding of the series, such as online discussion forums, fan fiction, and social media engagement. The feedback and reactions of viewers can also influence the direction of the series, as writers and producers may adjust the storyline or characters in response to audience feedback.

TV series are also influenced by broader cultural and social trends. They may reflect or challenge prevailing cultural norms and values and can be used as a tool for social commentary or political messaging. The broader cultural context in which a series is produced can also influence its reception by audiences, as different social groups may interpret the series in different ways depending on their own cultural and ideological perspectives. This is especially true for long-running series distinctive of broadcast television, which adopt a particular production model. Episodes are produced in parallel with the airing of the series, a few months in advance. The peculiar property of TV series being an open system interacting with the 'physical' world comes with another characteristic: a very low latency with respect to external stimuli and the capability of embedding in their narrative, almost immediately, events or facts from the real world. As an example, the Supreme Court ruling on *Roe vs. Wade*, which made abortion illegal in many U.S. states, can be cited. The ruling is dated June 24, 2022. The medical series *New Amsterdam* devoted an extremely critical episode to it, which aired on November 1, 2022. Considering the production and airing times, the reaction time to a major external event can be estimated to be a couple of months. Actually we believe that there is also a retro action in this interaction: changes in the narrative structure might impact on real cultural or social processes.

Generative AI

Another interesting direction that in a relative short future might have a huge impact on the television industry and media studies is the generation of new content using deep multimodal algorithms. A deep multimodal algorithm, as noted above, is an approach to machine learning that combines information from multiple sources, such as text, images, and audio. It is designed to take advantage of the complementary nature of different modalities to improve the accuracy and richness of the analysis and generation of artificial data.

Deep multimodal algorithms have become increasingly popular in recent years, particularly in the fields of NLP and computer vision. They are used for a wide range of tasks, from image and speech recognition to sentiment analysis and recommendation systems. At the core (learning phase) of most deep multimodal algorithms is the Transformer model, which was introduced by Vaswani and colleagues (2017). Transformers use the so-called self-attention mechanisms to encode and decode information in a very efficient way, through a neural network architecture that is capable of processing sequential data, such as text, as well as non-sequential data, such as images. It has been used successfully since then in a variety of natural language and computer vision tasks, often also in multimodal environment (Xu et al. 2022). Example of this include the Vision-and-Language pre-training models designed to learn joint representations of visual and textual information, enabling it to perform a range of multimodal tasks, such as image captioning and visual question answering (Li et al. 2020, Chen et al. 2023).

Generative methods, such as Generative Adversarial Networks (GANs) and Stable Diffusion, have gained immense popularity in recent years for their ability to create new content (Shahriar 2021). These multimodal generative algorithms can be used in a variety of applications such as image synthesis, speech synthesis, and NLP. For instance, in the field of image synthesis, multimodal generative algorithms can be used to generate photo-realistic images that do not exist in reality but look like they could be real. In speech synthesis, they can be used to generate human-like speech that is almost indistinguishable from the real thing. And in natural language processing, they can be used to generate text that is coherent and semantically meaningful.

GANs (Goodfellow et al. 2014) work by training two neural networks simultaneously: a generator and a discriminator. The generator creates new data by generating samples from a random noise input, while the discriminator evaluates the generated samples and decides whether they are real or fake. The two networks are then trained together in a process called adversarial training, where the generator learns to create more realistic samples by fooling the discriminator. Over time, the generator becomes better at generating realistic samples that can fool the discriminator, and the discriminator becomes better at identifying real from fake samples (Karras et al. 2018, Brock et al. 2019, Huang et al. 2021).

Stable Diffusion (Sohl-Dickstein et al. 2015, Ho et al. 2020), on the other hand, is a generative method that involves iteratively refining a noise

input to generate a sample that matches a given target distribution, empirically defined by real images. It uses the principles of diffusion processes in physics (in particular non-equilibrium statistical mechanics) to transform real images through steps of “small gaussian noise”, until all the information is lost, while a deep network learns how to invert the global (irreversible) process. This backward process is then used to generate artificial high-quality images from noise, creating artificial but realistic images of human faces, animals, and landscapes (Ho et al. 2020, 2021, Rombach et al. 2022). Furthermore, the attention mechanism embedded in the transformer can be used to ‘condition’ the generated image to be constrained to some given text, producing amazing artificial images (but also music) with a given style or specific content (see, for example: <https://stablediffusionweb.com/>).

GANs and diffusion models have shown great success in generating high-quality images and text in a multimodal setup. However, generating videos is more challenging due to the additional temporal dimension involved. In fact, to date, apart from a few exceptions, all applications for video exist in the so-called *deepfake*, i.e. the exchange of faces of a character in a movie scene, often porn. In order to generate videos, the models need to be able to capture the temporal dynamics of the scenes, for example how the objects and people in the video move and interact with each other over time. This requires incorporating the temporal dimension into the models, which can be done in several ways. The most mature approach is to use Recurrent Neural Networks (RNNs), such as LSTMs (Long Short-term Memory network), to model the temporal dependencies in the video frames. The RNNs can be trained to predict the next frame in the video sequence given the previous frames, and eventually combined with GAN architectures for interesting and promising applications to create complex but short duration videos (Clark et al. 2014, Vondrick et al. 2016, Munoz et al. 2020). Attempts have also applied diffusion methods to videos that forecast interesting future directions (Ho et al. 2022a, 2022b).

The TV series industry and media studies can in future gain from the development and consolidation of these generative methods, but in the short term the reverse is true: TV series could represent an ideal testing ground for the development of new deep multimodal algorithms for complex and long video analysis and generation. Developing multimodal deep algorithms suited to TV series data (for analysis or generative tasks) would also imply the creation of tools for a better exploration and understanding of certain facets of TV series’ complexity. As the technology advances and the data

becomes more diverse and abundant, we can expect to see more exciting and innovative ways of using these algorithms to enhance the TV viewing experience, as well as in production. One future application of generative multimodal deep learning algorithms in the realm of TV series could be, for example, in the creation of personalized interactive TV shows.

Overall, generative multimodal deep learning algorithms have the potential to revolutionize how we design, produce, create and consume TV shows, opening up new possibilities for personalization and interactivity. We cannot precisely predict the future of these generative AI algorithms as they apply to TV production, but change will come and, considering the accelerations of the last five years, mostly probably it will come sooner than expected. It is hard to predict the details of the impact on the TV industry and media studies, but there will be social, economic, and cultural impacts. Considering what is happening at time of writing (in April 2023) with “just one ChatBot” (<https://www.bbc.com/news/technology-65139406>, <https://www.wired.it/article/openai-blocca-chatgpt-in-italia/>), we could just about imagine what will happen when we can generate artificial media content of high complexity, long duration and realistic appearance.

BIBLIOGRAPHY

- Antelmi, Alessia, John Breslin and Karen Young (2018). “Understanding User Engagement with Entertainment Media: A Case Study of The Twitter Behaviour of Game of Thrones (GoT) Fans.” In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, 1-9. <https://doi.org/10.1109/GEM.2018.8516505>.
- Antonakaki, Despoina, Paraskevi Fragapoulou and Sotiris Ioannidis (2021). “A Survey of Twitter Research: Data Model, Graph Structure, Sentiment Analysis and Attacks.” *Expert Systems with Applications* 164, 114006. <https://doi.org/10.1016/j.eswa.2020.114006>.
- Avezzi, Giorgio (2017). “The Data Don’t Speak for Themselves: The Humanity of VOD Recommender Systems.” *Cinéma & Cie* 17(29): 51-66.
- Blei, David, Andrew Ng and Michael Jordan (2003). “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993-1022.
- Bost, Xavier, Serigne Gueye, Vincent Labatut, Martha Larson, Georges Linares, Damien Malinas and Raphaël Roth (2019). “Remembering Winter Was Coming: Character-Oriented Video Summaries of TV Series.” *Multimedia Tools and Applications* 78(24): 35373-99. <https://doi.org/10.1007/s11042-019-07969-4>.
- Bredin, Hervé and Grégory Gelly (2016). “Improving Speaker Diarization of TV Series Using Talking-Face Detection and Clustering.” In *Proceedings of the 24th ACM International Conference on Multimedia*, 157-61. <https://doi.org/10.1145/2964284.2967202>.
- Brock, Andrew, Jeff Donahue and Karen Simonyan (2019). “Large Scale GAN Training for High Fidelity Natural Image Synthesis.” In *International Conference on Learning Representation*, 1-29. <https://doi.org/10.48550/arXiv.1809.11096>.
- Burgess, Jean and Axel Bruns (2015). “Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research after the Computational Turn.” In *Compromised Data: From Social Media to Big Data*, edited by Greg Elmer, Ganaele Langlois and Joanna Redden, 93-111. New York: Bloomsbury Academic.
- Buschow, Christopher, Beate Schneider and Simon Ueberheide (2014). “Tweeting Television: Exploring Communication Activities on Twitter while Watching TV.” *Communications* 39(2): 129-149. <https://doi.org/10.1515/commun-2014-0009>.
- Chen, Fei-Long, Du-Zhen Zhang, Ming-Lun Han, Xiu-Yi Chen, Jing Shi, Shuang Xu and Bo Xu (2023). “VLP: A Survey on Vision-Language Pre-Training.” *Machine Intelligence Research* 20(1): 38-56. <https://doi.org/10.1007/s11633-022-1369-5>.

- Chen, Hsuan-Wei, Jin-Hau Kuo, Wei-Ta Chu and Ja-Ling Wu (2004). "Action Movies Segmentation and Summarization Based on Tempo Analysis." In *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 251-58. <https://doi.org/10.1145/1026711.1026752>.
- Clark, Aidan, Jeff Donahue and Karen Simonyan (2019). "Adversarial Video Generation on Complex Datasets." *arXiv:1907.06571*. <https://doi.org/10.48550/arXiv.1907.06571>.
- Crespo-Pereira, Verónica and Óscar Juanatey-Boga (2017). "Spanish TV Series on Twitter: What Social Media Audiences Say." In *Media and Metamedia Management*, edited by Francisco Campos Freire et al., 435-440. Cham: Springer.
- Cutting, James E., Jordan E. DeLong and Christine E. Nothelfer (2010). "Attention and the Evolution of Hollywood Film." *Psychological Science* 21(3):432-39. <https://doi.org/10.1177/0956797610361679>.
- Egger, Roman and Joanne Yu (2022). "Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts." *Frontiers in Sociology* 7. <https://doi.org/10.3389/fsoc.2022.886498>.
- Fanchi, Mariagrazia and Tarantino Matteo (2020). "Has Digital Transformation Impacted Gender Imbalance in Italian Cinema? A Data analysis of Creative Clusters 2004-2016." In *Digital Transformation in the Cultural and Creative Industries*, edited by Marta Massi, Marilena Vecco and Yi Lin. London: Routledge.
- Giachanou, Anastasia and Fabio Crestani (2016). "Like it or not: A Survey of Twitter Sentiment Analysis Methods." *ACM Computing Surveys (CSUR)* 2: 1-41. <https://dx.doi.org/10.1145/2938640>.
- Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio (2014). "Generative Adversarial." In *Proceedings of the International Conference on Neural Information Processing Systems*, 2672-2680. <https://doi.org/10.1145/3422622>.
- Grootendorst, Maarten (2022). "BERTopic: Neural topic modeling modelling with a class-based TF-IDF procedure." *arXiv arXiv:2203.05794v1*. <https://doi.org/10.48550/arXiv.2203.05794>.
- Guha, Tanaya, Naveen Kumar, Shrikanth S. Narayanan and Stacy L. Smith (2015). "Computationally deconstructing movie narratives: An informatics approach." In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2264-68. <https://doi.org/10.1109/ICASSP.2015.7178374>.
- Hagy, Madeleine, Ansley Hamilton and Caroline Miller (2019). "The Golden Age of Data, #ThisIsWe. A Content Analysis of Live Tweets of This Is Us." In *The Golden Age of Data. Media Analytics in Study & Practice*, edited by Don Grady. New York: Routledge.
- Hecking, Tobias, Vania Dimitrova, Antonija Mitrovic and H. Ulrich Hoppe (2017). "Using Network-Text Analysis to Characterise Learner Engagement in Active Video Watching." In *Proceedings of the 25th International Conference on Computers in Education*, edited by Chen W. et al., 326-335.
- Ho, Jonathan, Ajay Jain and Pieter Abbeel (2020). "Denosing Diffusion Probabilistic Models." *Advances in Neural Information Processing Systems* 33: 6840-6851. <https://arxiv.org/abs/2006.11239>.

- Ho, Jonathan, Chitwan Saharia, William Chan, David J. Fleet, Mohammad Norouzi and Tim Salimans (2022). "Cascaded Diffusion Models for High Fidelity Image Generation." *Journal of Machine Learning Research* 23(47): 1-33. *arXiv* <http://arxiv.org/abs/2106.15282>.
- Ho, Jonathan, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi and David J. Fleet (2022). "Video Diffusion Models." *arXiv* <http://arxiv.org/abs/2204.03458>.
- Ho, Jonathan, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet and Tim Salimans (2022). "Imagen Video: High Definition Video Generation with Diffusion Models." *arXiv* <https://arxiv.org/abs/2210.02303>.
- Huang, Qingqiu, Yuanjun Xiong, Yu Xiong, Yuqi Zhang and Dahua Lin (2018). "From Trailers to Storylines: An Efficient Way to Learn from Movies." *arXiv* <https://arxiv.org/abs/1806.05341>.
- Huang, Zhizhong, Shouzhen Chen, Junping Zhang and Hongming Shan (2021). "PFA-GAN: Progressive Face Aging with Generative Adversarial Network." In *IEEE Transactions on Information Forensics and Security* 16: 2031-45. <https://doi.org/10.1109/TIFS.2020.3047753>.
- Ibrahim, Ania, Ahmed Elbagoury, Mohamed S. Kamel and Fakhri Karray (2018). "Tools and Approaches for Topic Detection from Twitter Streams: Survey." *Knowledge and Information Systems* 3: 511-539. <https://doi.org/10.1007/s10115-017-1081-x>.
- Innocenti, Veronica and Guglielmo Pescatore (2018). "The Evolution of Characters in TV Series: Morphology, Selection, and Remarkable Cases in Narrative Ecosystems." In *Reading Contemporary Television Universes. A Narrative Ecosystem Framework*, edited by Ilaria A. De Pascalis and Paola Brembilla. New York and London: Routledge.
- Karras, Tero, Timo Aila, Samuli Laine and Jaakko Lehtinen (2018). "Progressive Growing of GANs for Improved Quality, Stability, and Variation." *arXiv* <http://arxiv.org/abs/1710.10196>.
- Lee, Daniel and H. Sebastian Seung (1999). "Learning the Parts of Objects by Non-negative Matrix Factorization." *Nature* 401: 788-791. <https://doi.org/10.1038/44565>
- Li, Linjie, Yen-Chun Chen, Yu Cheng, Zhe Gan, Licheng Yu and Jingjing Liu (2020). "HERO: Hierarchical Encoder for Video+Language Omni-Representation Pre-Training." *arXiv* <http://arxiv.org/abs/2005.00200>
- Maidi, Madjid, Veronica Scurtu and Marius Preda (2011). "Characters Identification in TV Series." In *2011 IEEE International Symposium on Multimedia*, 143-48. <https://doi.org/10.1109/ISM.2011.31>.
- McCallum, Shiona (2023). "ChatGPT banned in Italy over privacy concerns." *BBC NEWS*
- Miech, Antoine, Dimitri Zhukov, Jean-Baptiste Alayrac, Makarand Tapaswi, Ivan Laptev and Josef Sivic (2019). "HowTo100M: Learning a Text-Video Embedding by Watching Hundred Million Narrated Video Clips." In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2630-40. <https://doi.org/10.1109/ICCV.2019.00272>.
- Molteni, Luca and Joaquin Ponce de Leon (2016). "Forecasting with Twitter Data: An Application to USA Tv Series Audience." *International Journal of Design & Nature and Ecodynamics* 3: 220-29. <https://doi.org/10.2495/DNE-V11-N3-220-229>.
- Mottaghinia, Zahib, Mohammad-Reza Feizi- Derakhshi, Leili Farzinbash and Pedram Salehpour (2021). "A Review of Approaches for Topic Detection in

- Twitter.” *Journal of Experimental & Theoretical Artificial Intelligence* 5: 747-773. [10.1080/0952813X.2020.1785019](https://doi.org/10.1080/0952813X.2020.1785019)
- Munoz, Andres, Mohammadreza Zolfaghari, Max Argus, and Thomas Brox (2021). “Temporal Shift GAN for Large Scale Video Generation.” In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3179-3188. *arXiv* <http://arxiv.org/abs/2004.01823>.
- Nagrani, Arsha and Andrew Zisserman (2018). “From Benedict Cumberbatch to Sherlock Holmes: Character Identification in TV Series without a Script.” *arXiv* <http://arxiv.org/abs/1801.10442>.
- Naldi, Maurizio and Paola Dalla Torre (2022). “Uncovering the Narrative Structure of “Breaking Bad” Through a Multi-Dimensional Quantitative Analysis.” *SERIES – International Journal of TV Serial Narratives* 8(1): 17-32. <https://doi.org/10.6092/issn.2421-454X/14503>
- Park, Tae Jin, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J. Han, Shinji Watanabe and Shrikanth Narayanan (2022). “A Review of Speaker Diarization: Recent Advances with Deep Learning.” *Computer Speech & Language* 72, 101317. *arXiv* <http://arxiv.org/abs/2101.09624>
- Pehlivan, Zeynep (2021). “Linking Twitter Archives with Television Archives.” In *The Past Web*, edited by Daniel Gomes, Elena Demidova, Jane Winters and Thomas Risse. Cham: Springer. https://doi.org/10.1007/978-3-030-63291-5_10.
- Pescatore, Guglielmo (ed.) (2018). *Ecosistemi Narrativi. Dal fumetto alla serie tv*. Roma: Carocci.
- Pescatore, Guglielmo and Marta Rocchi (2019). “Narration in Medica Dramas. Interpretative Hypotheses and Research Perspectives.” *La Valle dell’Eden* 34: 107-115.
- Pescatore, Guglielmo, Veronica Innocenti and Paola Brembilla (2014). “Selection and Evolution in Narrative Ecosystems. A Theoretical Framework for Narrative Prediction.” In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) IEEE Computer Society*, 1-6. <https://doi.org/10.1109/ICMEW.2014.6890658>.
- Pilař, Ladislav, Lucie Kvasničková Stanislavská, Roman Kvasnička, Petr Bouda and Jana Pitrová (2021). “Framework for Social Media Analysis Based on Hashtag Research.” *Applied Sciences* 11(8): 3697. <https://doi.org/10.3390/app11083697>.
- Rao, Anyi, Linning Xu, Yu Xiong, Guodong Xu, Qingqiu Huang, Bolei Zhou and Dahua Lin (2020). “A Local-to-global Approach to Multi-modal Movie Scene Segmentation.” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10146-10155. *arXiv* <http://arxiv.org/abs/2004.02678>.
- Rocchi, Marta (2022a). “Applying Automatic Text Analysis Methodologies to Audiovisual Serial Product.” *Cinergie—Il Cinema e le altre Arti* 22: 173-187. <https://doi.org/10.3390/app11083697>.
- Rocchi, Marta (2022b). “Environmental Misinformation and Audiovisual Serial Narratives. An Automatic Analysis of the Twitter Social Discursiveness on ‘Seaspiracy’.” *Elephant & Castle* 28: 124-134.
- Rocchi, Marta and Allegra Sonogo (2022). “Tribunali e giustizia nelle narrazioni seriali del prime-time statunitense.” *Griseldaonline* 21(2): 79-93. <https://doi.org/10.6092/issn.1721-4777/15434>.

- Rocchi, Marta and Guglielmo Pescatore (2022). "Modeling Narrative Features in TV Series: Coding and Clustering Analysis." *Humanities and Social Sciences Communications* 9(333). <https://doi.org/10.1057/s41599-022-01352-9>
- Rocchi, Marta and Martina Anselmeti (2021). "L'evoluzione del legal drama, stereotipi di genere e il caso di For the People (ABC 2018-2019)." *Ocula* 22(25): 229-245. <https://doi.org/10.12977/ocula2021-16>.
- Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser and Björn Ommer (2022). "High-resolution Image Synthesis with Latent Diffusion Models." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684-10695. *arXiv* <http://arxiv.org/abs/2112.10752>.
- Ruffino, Marco and Paola Brembilla (2016). "Narrative Ecosystems Through the Network Analysis Lens. Step One: The Production of U.S. TV Series, Between Capital and Labor Strategies." *SERIES – International Journal of TV Serial Narratives* 2(1): 55-68. <https://doi.org/10.6092/issn.2421-454X/6164>.
- Sala, Teresa M. and Mariona Bruzzo (eds) (2019). *I-MEDIA-CITIES Innovative e-environment for Research on Cities and the Media*. Barcellona: Edicions de la Universitat de Barcelona.
- Scharl, Arno, Alexander Hubmann-Haidvogel, Alistair Jones, Daniel Fischl, Ruslan Kamolov, Albert Weichselbraun and Walter Rafelsberger (2016). "Analyzing the Public Discourse on Works of Fiction-Detection and Visualization of Emotion in online Coverage about HBO's Game of Thrones." *Information Processing & Management* 52(1): 129-138. <https://doi.org/10.1016/j.ipm.2015.02.003>.
- Schrage, Michael (2020). *Recommendation Engines*. Boston: Mit Press.
- Shahriar, Sakib (2021). "GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network." *arXiv* <http://arxiv.org/abs/2108.03857>.
- Skowron, Marcin, Martin Trapp, Sabine Payr and Robert Trapp (2016). "Automatic Identification of Character Types from Film Dialogs." *Applied Artificial Intelligence* 30(10): 942-73. <https://doi.org/10.1080/08839514.2017.1289311>.
- Sohl-Dickstein, Jascha, Eric A. Weiss, Niru Maheswaranathan and Surya Ganguli (2015). "Deep Unsupervised Learning using Nonequilibrium Thermodynamics." In *International Conference on Machine Learning*, 2256-2265. PMLR. *arXiv* <http://arxiv.org/abs/1503.03585>.
- Tapaswi, Makarand, Martin Bauml and Rainer Stiefelbogen (2012). "Knock! Knock! Who Is It; Probabilistic Person Identification in TV-Series." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2658-65. <https://doi.org/10.1109/CVPR.2012.6247986>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin (2017). "Attention is All you Need." In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000-10.
- Vondrick, Carl, Hamed Pirsiavash and Antonio Torralba (2016). "Generating Videos with Scene Dynamics." *Advances in Neural Information Processing Systems* 29. *arXiv* <http://arxiv.org/abs/1609.02612>.
- Williams, Apryl and Vanessa Gonlin (2017). "I Got All My Sisters with Me (on Black

- Twitter): Second Screening of How to Get Away with Murder as a Discourse on Black Womanhood.” *Information, Communication & Society* 20(7): 984-1004. <https://doi.org/10.1080/1369118X.2017.1303077>.
- Xu, Peng, Xiatian Zhu and David A. Clifton (2022). “Multimodal Learning with Transformers: A Survey.” *arXiv* <https://doi.org/10.48550/arXiv.2206.06488>.
- Zhang, Yi-Fan, Changsheng Xu and Hanqing Lu (2008). “Automatic Character Identification in Feature-length Films.” In *2008 IEEE International Conference on Multimedia and Expo*, 1469–72. <https://doi.org/10.1109/ICME.2008.4607723>.
- Zimbra, David, Ahmed Abbasi, Daniel Zeng and Hsinchun Chen (2018). “The State-of-the-art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation.” *ACM Transactions on Management Information Systems (TMIS)* 2: 1-29. <https://doi.org/10.1145/3185045>
- Zorloni, Luca (2023). “OpenAi blocca ChatGPT in Italia” *Wired*. <https://www.wired.it/article/openai-blocca-chatgpt-in-italia/> (last accessed 20-04-23).

EXPLORING TV SERIALITY AND TELEVISION STUDIES THROUGH DATA-DRIVEN APPROACHES



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.ec022085>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

3. Looking for Lexical Signatures in *Gomorrah*

Paola Dalla Torre, Paolo Fantozzi,
Maurizio Naldi and Placido Pellegriti

◀ ABSTRACT

Gomorrah is an Italian crime drama TV series that has been appreciated worldwide, being sold in 190 countries, despite its wide use of Neapolitan dialect, hardly understandable without the use of subtitles even for most Italians. Scholars immediately approached the study of this serial phenomenon, analysing it from different points of view, framing it within the broader context of studies on the new Italian television and its serial products. Our approach to *Gomorrah*, taking these elements into account, adds a new perspective that concerns character recognition, an emerging branch of research, to associate dialogues with characters and identify the verbal features of characters. We have then chosen *Gomorrah* as a challenging dataset to perform character recognition. We rely on the transcripts of the series after a pre-processing stage to standardize the lexicon despite the vagaries of dialect and remove stopwords. A machine learning approach, based on a selection of tools, is then employed to identify characters from the lexicon they employ. The problem is approached as a multi-class classification scheme. We compare several representations of texts, including the simple one-hot encoding and more advanced embedding techniques. The results are presented through a confusion matrix, which can also serve to identify similarities in the linguistic profiles of characters.

KEYWORDS

Character recognition; *Gomorrah*; machine learning; text analysis.

Introduction

As early as 2005, Moretti spoke about the need to apply quantitative methods to literary works to give birth to *distant reading* that would take the place of the close reading method used until then. By *distant reading*, he meant a method involving moving away from the text and observing it from a distance. That distance would not be an obstacle to knowledge but rather a specific form of it: “distance makes one see details less, true: but it makes one understand relationships, patterns, forms better”. Distant reading allows us to move from text to pattern or rather to patterns (Moretti 2005: 3).

The book by Moretti represented a seminal work for introducing quantitative methods in the analysis of literature. But the advent of new media and new narrative forms makes a case for the extension and development of those techniques to them. This is particularly true for TV series, which represent a dominant form in the present landscape of fiction works (Hammond 2005).

As a further step in the introduction of quantitative methods in the analysis of TV series, this paper reports some early results of our work on the linguistic corpus of the *Gomorra*, aiming to offer new models of interpretation of television series, following the path of digital humanities and, more specifically, the methods used by Franco Moretti in the literature.

Our approach follows the reflections made by Moretti himself a few years after his seminal work that we have already mentioned. In fact, Moretti himself observed that while it was true that quantitative methods had established themselves in the study of literature, his idea of “abstract models” had not found wide acceptance or yielded the expected results. It was necessary to restart after that “false movement” and integrate new techniques such as text mining, topic modeling, content analysis, and sentiment analysis with twentieth-century literary theories such as hermeneutics or history (Moretti 2022). However, Moretti recognizes that the digital humanities risk losing

sight of form, essence, and social dimension owing to the overwhelming presence of data. In his own words, “Interpretation moves between form and the world, going in search of the historical meaning of works; quantification moves between form and form, attempting to trace the coordinates of a future atlas of literature” (Moretti 2022: 32). A synthesis between those two contrasting movements has to be found as nothing of the sort has yet been fully achieved, and we must keep trying.

In this paper, we apply text-mining techniques to the corpus made of the dialogues in *Gomorra*, integrating media-studies interpretation and linguistic analysis in the footsteps of what linguists have already been conducting on the linguistic corpus of television series for several years (Bednarek 2018). Our final aim is to uncover the underlying association between a character and his lexical signature.

Gomorra

Gomorra has been a huge national and international success. In America, it has even been compared to major crime series such as *The Sopranos* or *The Wire*, and in 2016 the New York Times ranked the series at No. 3 of best series of the year and in 2019 again, the New York Times ranked it at No. 5 among the best international (non-U.S.) productions of the last ten years. In Italy, it has broken all TV rating records and has become a widespread cultural phenomenon, an object of cult, spurring, however, controversies related to the depiction of the figure of the criminal and its glamorous rendering, as is often the case in series devoted to the crime genre.

Born from the idea of Roberto Saviano, to whom we owe the investigative book of departure, the series arrives after the movie directed by Matteo Garrone (2008) and the theatrical show. It was produced by Cattleya, Fandango and Sky from 2014 to 2021 and is structured into five seasons for a total of 58 episodes. The main directors were Stefano Sollima, Francesca Comencini, and Claudio Cupellini, while the main scriptwriters were Leonardo Fasoli, Stefano Bises, and Maddalena Ravagli. Stefano Sollima, the showrunner of the series, is also the showrunner of *Romanzo criminale*, the first serial production of the Sky model, the pay TV that revolutionized the audiovisual environment for the small screen in the 2000s, opening the Third Golden Age of TV in our country as well (Bernardelli 2016). Sky has invested in producing content with uncomfortable topics and high-quality standards. Following the model of great American series, to which audiences are becom-

ing increasingly familiar, Sky has begun to produce series with well-defined genres, characterized by multi-seasonality, with horizontal and vertical storylines, following the serialized series model defined by Mittel. Such series are also careful to interact with their audiences, giving much space to fandom (Napoli and Tirino 2016). Barra and Scaglioni write that the Sky editorial policy is also to tell about “bad guys” (Barra and Scaglioni 2015) on par with the “difficult man” of American quality television (Martin 2013).

The story is well-known and focuses on a Camorra clan (the Savastanos) fighting for control of the area and various illegal trafficking and on the rivalry of the two former friends, Ciro and Genny, son of the boss. The two of them are the real protagonists, present in all seasons, alongside several other characters who rotate in and out of the extended narrative universe of the series. Its characters are divided into very specific hierarchies: the bosses, the sidekicks, women, law enforcement, and the world of finance and business. The topics narrated, in fact, are not only those related to crime, the Camorra, drugs, and power but also those of family, friendship, and love. Public and private interact in a mixture of realistic drama and melodrama. The style is raw, direct, and violent, and for the performers, little-known actors (who became famous later) capable of using dialect were chosen. Opting for crude spoken language also sought strong adherence to reality. Adherence to reality is also visible in the choice of locations: the district of the sails of Scampia as well as the real places of the criminal Naples that Saviano had already described in his books and Garrone had narrated in his film.

Literature Review

Gomorra has received much attention in the TV series criticism context. Scholars have analyzed it right after its appearance, adopting different points of view and different methodological approaches. It has been a global hit and is considered on a par with the great American quality TV series (McCabe and Akass 2007). It is certainly the best result of the new standards of products made by Italian television (Scaglioni and Barra 2013).

In this section, we review the most relevant essays in Media Studies concerning the series and the most recent results in the quantitative analysis of TV series. The first line of research we analyze concerns the transmediality of *Gomorra* and its huge success both as a narrative ecosystem (De Pascalis 2018) and as a global brand (Benvenuti 2017). The series, in fact, is only

the third remediation, as Grusin would say (Bolter and Grusin 1999), of the original text consisting of Roberto Saviano's novel (already a collation of previous reports). A play was derived from this novel, followed by the film directed by Matteo Garrone and, finally, the series. This movement across media may be dubbed "media colonization" (Guerra et al. 2017). However, it does not result from a reasoned expansion but rather through random and irregular choices, according to a weak transmediality, where projects accumulate only by a logic of simple reference to the source novel.

This research line also includes studies concerning fandom and the public's creativity in the resemantization of the series with the making of memes and videos. *Gomorra* is an expanded universe also because of the audience: as described in (Napoli and Tirino 2016: 108), a symbolic pulp space has been built with the contributions of the public, the platform, and viral videos, where elements borrowed from gangster and spaghetti western movies mix with the cult expressions of *Gomorra*. Videos mainly adopt a parody and quasi-mythological tone with inter-media links and often resort to memetic techniques (Chen 2012, Dawkins 2006) with remix culture elements (Manovich 2007).

Another line of research has paid attention to the narrative structure and characters, e.g., the analysis of female characters in a story where male characters dominate (Milkowska-Samul 2017) or the analysis of main characters as anti-heroes (Martin 2017, Russo 2017, Bernardelli 2016). As to the latter issue, the process of viewers' identification with a morally repugnant protagonist has been described by Russo with the following words:

Ciro Di Marzio, Genny and the other primary characters of *Gomorra* are quintessentially modern inasmuch as they function in the narrative as heroes with whom we align and to whom we temporarily pledge our allegiance, only to eventually be disappointed and repulsed by them. William Brown called on Stanley Cavell's notion of revulsion to suggest an 'ethical mode of engagement' with explicit violence in extreme films (Russo 2017: 17).

Many analyses have focused on the problem of realism in the series, relating it to the original text. The series has been accused of too much realism and excessive violence. Instead, most studies observe that the series' structure is not so realistic but is often closer to other narrative types. For example, Bernardelli highlighted how the series moves between realistic content and melodramatic fictional representation (Bernardelli 2017). Applying the

meaning that Mittel assigns to melodrama, defined as a style of excess and exaggeration of feelings, actions, and words, aimed at narrating feelings and creating pathos, Bernardelli states that contemporary TV series are dominated by anti-heroes who perform deplorable actions and with whom it would be impossible to connect if there were not this very melodramatic mechanism. *Gomorrah*, therefore, has a realistic setting for its raw and direct depiction, bringing to life almost chronicle-like scenes from the lives of the bosses, but, at the same time, it takes on the melodramatic style to make its narrative acceptable to the viewer. In his own words:

The rawness of the events described, the realism, is constantly balanced by strongly theatrical, melodramatic, and often emphatic acting. It is like saying that if you have the viewer bear the weight of reality, you must then emphasize the fictionality of what you represent through the melodrama exhibited by acting: a realistic substance corresponds to a melodramatic form (Bernardelli 2017: 9).

Pescatore and Rocchi (2020) also mentioned a particular “realism formula” for *Gomorrah*, which is closer to the structure of a narrative genre than a documentary one. They analyzed the lexical corpus of Saviano’s book and the lexical corpus of the synopses of the episodes to extract latent topics, that is, lexical/semantic regularities, and find the structure of the series to be related to the gangster genre and its stereotypes. The realism of *Gomorrah* is only conventional because all the stereotypes associated with the gangster genre, such as the mafia crime family, the relationship between generations, and the mechanism of drug dealing, clearly emerge from the study (Pescatore and Rocchi 2020: 105).

The issue of *Gomorrah*’s realism is also investigated by Variano (2019), who defines the series’ language as “furbesco seriale”: a combination of realistic dialectal slang and a reinvention by the scriptwriters, aimed at making the dialogue comprehensible (Variano 2019). The language of *Gomorrah* moves between standard Italian and dialectal Italian.

As with all the ingredients in the series, the language has also conformed to the choice of realism. The Gomorrese is the Neapolitan spoken in Scampia and Secondigliano, softened by some simplifying interventions (made in dubbing in some cases) to make the language as understandable as possible without taking away authenticity to the dialogues (Variano 2019: 287).

Therefore, the series ends up exhibiting a language that is realistic but reinvented by screenwriters at the same time.

The resulting mixture of languages is difficult to translate. Some studies have focused on the difficulties of translating dialogues in English for subtitles (Fruttaldo 2018, Raffi 2017). However, Fruttaldo has pointed out how *Gomorra*'s dialogues can be translated in subtitles and how cross-cultural remediation is possible, stating that

the (re)translation process, when it comes to works that are representative of an extremely local society, may in a way impoverish the source context references in the target language environment; this was the case for *Gomorra*, where a lower culture-specificity can be highlighted in the translation process (Fruttaldo 2018: 156).

If we move to the quantitative analysis of TV series, despite the seminal work by Moretti on fiction works, dating nearly a decade (Moretti 2005), the studies are relatively recent. Most works have concentrated on applying social network analysis techniques (the resulting network is aptly named character network) based on dialogues or other forms of interaction among the characters to describe the dynamics of the character ecosystem and its power structure. Labatut and Bost have surveyed the tools employed for that purpose in (Labatut and Bost 2019). The tools do not differ significantly from what can be employed for novels and other works of literary fiction. Both Edwards et al. and Vala et al. (the latter with specific reference to literary works) have highlighted the difficulties of the task in (Edwards et al. 2018, Vala et al. 2015). Specific examples of the application of social network analysis to TV series are the analysis of the *Game of Thrones* performed by Beveridge and Shan (2016), and the analysis of the narration of *Breaking Bad*, again *Game of Thrones*, and *House of Cards* (Bost et al. 2018). Chao et al. (2018) have employed a two-mode network to analyse the dynamics of character activities and plots in movies. Dialogues have been examined in (Fronzetti Colladon and Naldi 2018) to analyse the dominance of characters in the plot (or, vice versa, the choral features of the narration) regarding the *Big Bang Theory* series. A more extensive analysis, including the concentration analysis of video presence and locations in addition to dialogues has been carried out by Naldi and Dalla Torre (2022) for *Breaking Bad*. Also, Fronzetti Colladon and Naldi (2019) have analysed the relationship between the text and the series' success, again in the *Big Bang Theory*. Social networks have also been employed to validate literary theories, e.g.,

about the relationship between the number of characters and the settings (Jayannavar et al. 2015). Waumans et al. carried out an even more ambitious plan of obtaining the signature of a novel’s story through the topological analysis of its character network (Waumans et al. 2015).

Our analysis of *Gomorra* is lexically based and uses subtitles as its source, observing that dialectal expressions are often not amenable to consistent lemmatization, and there are significant differences between scripts and the actual dialogues. We avoid the language dilemma by excerpting dialogues from subtitles and then performing lemmatization. Since subtitles exploit a softened form of the dialect that mimics the Italian language, we do not have to consider pure dialect expressions at this stage. We use an approach that Franco Moretti has called “Quantitative Formalism”: “take an aesthetic form and disassemble it down to its basic elements: transforming – or rather, let’s say, reducing – a novel to its paragraphs, or a drama to a sequence of linguistic exchanges” (Moretti 2022). In our case, we disassemble by reducing dialogues to their words, adopting the bag-of-words approach. Our final aim is to uncover the lexical signature of each character in *Gomorra*. As long as the lexicon is not distorted (e.g., flattened) in the spoken language-to-subtitles process (i.e., words that are different in the spoken dialogues stay different in their subtitle version), using subtitles does not impact lexical signatures.

Our main research question is then: can we identify a *Gomorra* character from what he/she says? Can we find a lexical signature in *Gomorra*? To do this, we apply a perspective that concerns character identification, an emerging branch of research, to associate dialogues with characters and identify the lexical features of characters.

Methodology and Results

We have applied several quantitative techniques to help us build a profile of each character and identify them based on their lexicon only. In this section, we describe the features we have analysed so far and the methods we have employed. The results of this section are obtained based on the dialogues of the whole series as transcribed in the subtitles.

The features we are interested most in concern both the lexicon employed and its use by the characters. Our underlying research hypothesis is that each character bears a unique lexicon signature, i.e., a personal set of

words that he/she employs most. If this hypothesis holds true, assigning a text extracted from the dialogues to a specific character could be possible.

In order to progress towards that ultimate goal, we have so far analysed the following features:

- Word presence;
- Lexicon similarity and variety;
- Character presence;
- Character identification.

Word Presence

We first examine the lexicon employed in *Gomorrah*. A lexicon is made of words. The first thing we wish to observe is the set of words employed in the series, their overall quantity and frequency.

We start with observing the overall number of words uttered during each episode. In Figure 1, we show the time series of the number of words per minute. Using the density of words removes the bias due to the length of episodes. In fact, episodes differ widely in length, ranging from 35 to 55 minutes. The average number of words per minute is 47 over the whole series. If we take a closer look at the season, that average is 46 minutes for Seasons 1 and 2, then lowers to 43 and 41 minutes, respectively, in Seasons 3 and 4, to grow to 45 minutes in the last season. Is that average figure low or high? It should be compared with that of other series in the same genre. Of course, as we can see from the graph, we also have variations from epi-

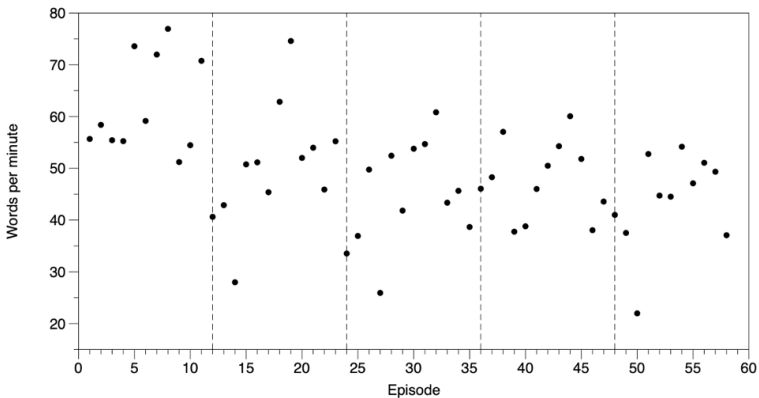


FIGURE 1
Overall density of words throughout the series.



FIGURE 2
Word cloud observed in *Gomorrah*.

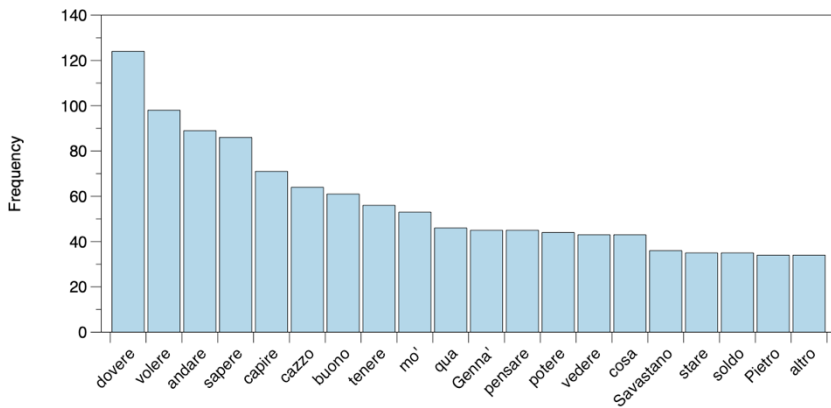
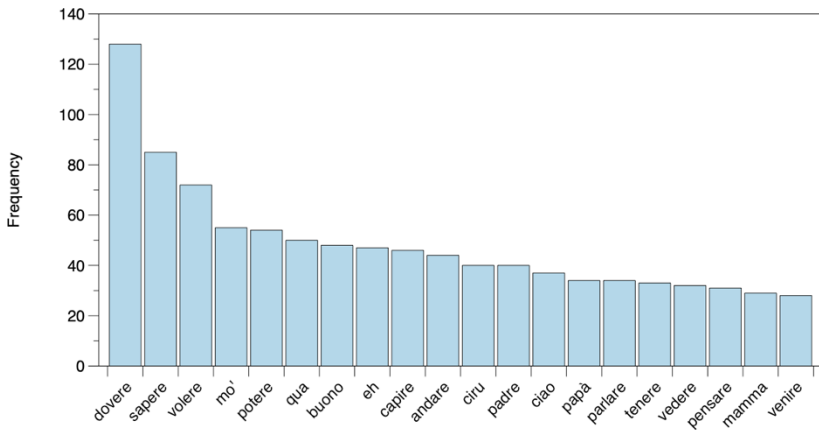
sode to episode and among seasons. In most cases, the average word density is between 40 and 60. The variations are related to the difference in the action/dialogue content in the episode, and it would also be interesting to see if they are related to the different screenwriters.

Though an ordered list could represent this, we prefer a more intuitive pictorial representation through a word cloud. The words in the series are shown in a horizontal or vertical arrangement, with a size proportional to their frequency. Here we do not pursue the many possibilities of text exploration through word clouds as Heimerl et al. (2014) proposed.

We show the word cloud in Figure 2, where all the words have been lemmatized: nouns have been reduced to their singular forms, and verbs have been reduced to their infinitive form. The largest words (i.e., the most frequently used ones) show us some interesting things. Bad language is certainly a dominant feature of the language in the series, and we understand well why. Variano's linguistic study also highlights the frequency of the use of insulting epithets and trivialisms in *Gomorrah's* dialogues, which are very often used with a semantic extension so that, for example, the word *cazzo* (*dick*) that we see in large print in our word cloud, comes to mean *niente*,

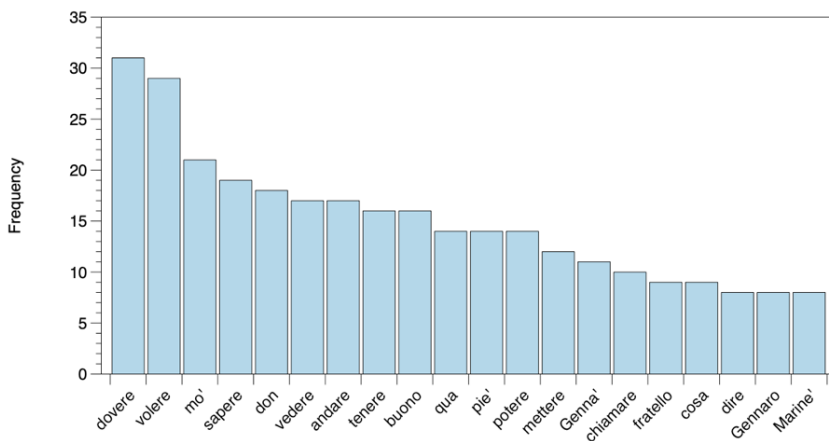
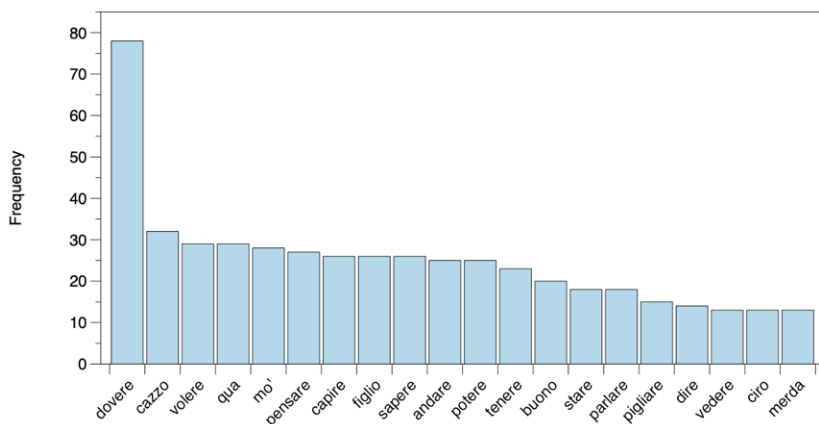
nessuno but also *qualcosa*, or even *interessarsi*, *preoccuparsi di qualcuno* or also *braccare* (Variano 2019: 296).

On closer inspection, two verbs emerge. On the one hand, we see the verb *dovere* (*to must*): as if the whole universe of *Gomorra* were dominated by unwritten rules that everyone must obey. On the other hand, we observe the verb *volere* (*to want*), which corresponds to the protagonists' aspirations, and their quest for power, money, and supremacy. The other strong element that emerges is the topic of family: *padre* (*father*), *mamma* (*mother*), *figlio* (*son*), *famiglia* (*family*) appear as dominant in the cloud.



FIGURES 3, 4
Lexicon employed by Genny (Gennaro Savastano) and Lexicon employed by Ciro.

Gomorrhah runs on this triad: duty, will, and family. And it does according to a mode that follows genre conventions, especially for gangsters, with its stereotyped elements and character typecasting. So, the first lexical signature we can identify in *Gomorrhah* is a stereotyped narrative, as other scholars have already said (Pescatore and Rocchi 2020).



FIGURES 5, 6
Lexicon employed by Pietro and Lexicon employed by Patrizia.

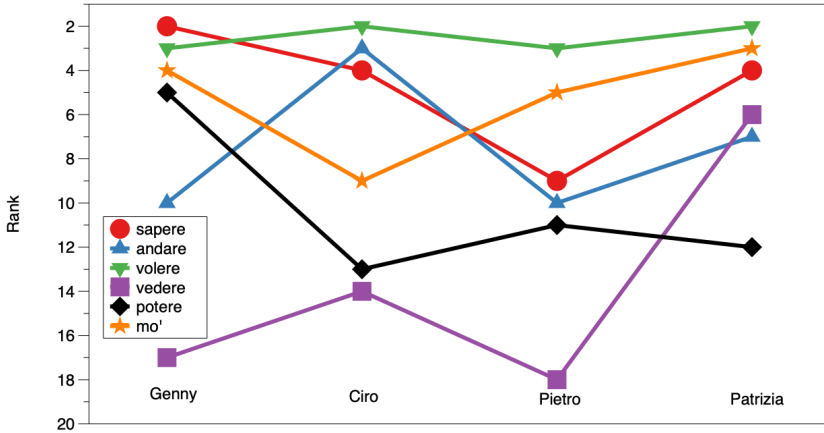


FIGURE 7
Word ranking comparison.

Lexicon Similarity and Variety

We can now dig deeper and examine the specific lexicon employed by the main characters in a comparative fashion. In Figures 3 through 6, we show the words mostly employed by four major characters; the old boss Pietro Savastano, his son Genny (Gennaro Savastano), the once-aid and then rival Ciro, and Patrizia, who starts her criminal career as a liaison officer for Pietro to later become the sole boss of Secondigliano.

As already observed for the whole series, the word *dovere* (to must) ranks first for all those characters. In Figure 7, we have reported the ranks of some words for the four characters. We see that *volere* (to want) ranks second for Ciro and Patrizia, and third for Genny and Pietro. Also, a common interjection (*mo*) ranks quite high for all the characters. If we compare the most used words by Ciro and Genny, the two bosses, we can spot many similarities. The vocabulary is almost the same: they use the same phraseology, as if they share a linguistic code that is that of the bosses, of those in charge. If we compare Figures 3 and 5, we see that Pietro and Genny (father and son) share 14 words in their Top 20.

The similarity between the characters' lexicons can be shown quantitatively for the whole set of characters if we compute the Jaccard Index between any two characters. The Jaccard Index is the ratio of the number of common words between the two characters and the overall number of words

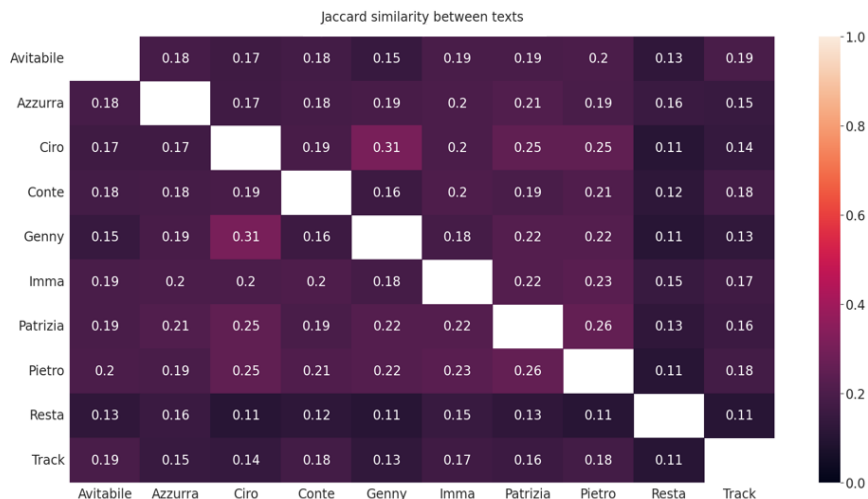


FIGURE 8
Jaccard Index.

they employ and measures the similarity between the lexicons employed by two characters and takes values in the $[0,1]$ range. If that index is close to 0, the two characters use nearly disjoint sets of words. On the opposite end of the range, a Jaccard Index close to 1 shows a nearly identical vocabulary. In Figure 8, we show the Jaccard Index in a matrix form to compare the most important characters. We see that it takes the highest value (0.31) for the Genny-Ciro pair (i.e., they roughly share 30% of their vocabulary), two emerging bosses belonging to the same generation. Slightly lower values show up between Pietro and Ciro (0.25), and Pietro and his son (0.22). On the other end of the spectrum, we find Resta, whose lexicon exhibits very low similarity values with all the other characters (0.11 with Pietro, Ciro, and Genny). Resta is a white collar, who, though joining Genny in a business endeavour, comes from a very different background, being a successful manager. It is to be noted that the Jaccard Index does not account for the frequency of the words, but just for their presence, in a binary fashion. For that reason, we observe a Jaccard Index of 0.22 between Genny and Pietro, while they share 14 out of their Top 20 words (i.e., a commonality ratio of 70%). Probably, they do not share many of their least frequent words.

This linguistic bosses' code, this bosses' signature, frequently uses verbs like must and want and, also, many words concerning the family (father,

son, mother) or proper names of characters linked to the Savastano's family or otherwise to bosses. So, we have exactly the three corners of the triangle delimiting the lexical morphology of *Gomorra*. We can also note another feature: the use of interjections, trivialisms, slang and colloquial phraseology that make up the low and lower-middle register of the characters. In Figure 7, we have noted the widespread presence of the interjection *mo'*. We also observe bad language in the male characters' lexicon. We observe *cazzo* (*dick*) and *merda* (*shit*) in ranks 2 and 20, respectively, for Pietro, with *cazzo* appearing in rank 6 for Ciro. So the stereotyped sign is combined with the cadence of "furbesco seriale", which makes *Gomorra*'s lexicon highly expressive and endowed with great representative force (Variano 2019: 296).

We can conclude that in *Gomorra* we have a bosses' signature: a code of words that all the bosses share and use, a code that marks the general linguistic nature of the series. As noted by Variano (2019: 295), the lingo of *Gomorra* is not the spoken language in Naples or its outskirts but rather a language used for a purpose in certain situations.

After examining the similarity of the lexica employed by different characters, which may serve as a discriminating feature to identify each character by his/her lexicon, we now consider lexicon variety. We are back to a characteristic of each character, though we can, of course, compare how the different characters perform. By variety we mean the richness of their vocabulary, how many words they employ and how often they keep em-

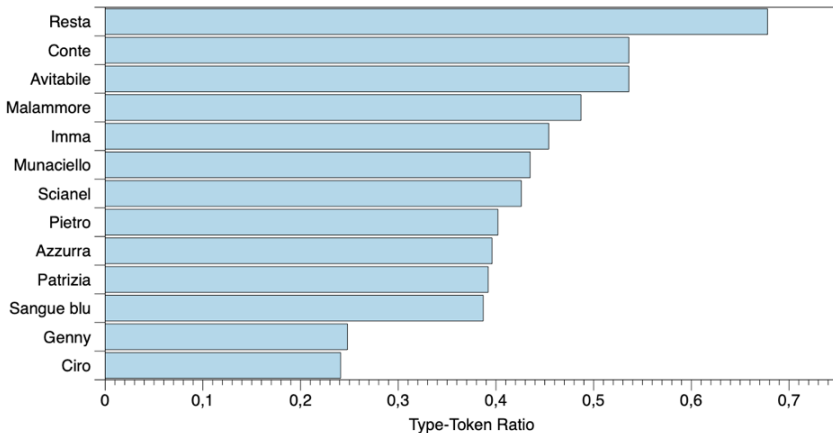


FIGURE 9
Type-Token Ratio.

ploying the same words repeatedly. We measure variety by a single indicator: the Type-Token Ratio (TTR), i.e., the ratio between the number of unique words and the overall number of tokens. Values closer to 1 imply a rich language, while values closer to 0 show a poor vocabulary. In Figure 9, we report the TTR for the major characters. We observe the poor language used by *Ciro* and *Genny*, who exhibit a TTR of around 0.25, while *Resta*, a white-collar character, reaches the highest value here. *Conte* and *Avitabile* are the two runner-ups, respectively, who play the role of a somewhat mannered boss and a borderline boss who runs a hotel aside from his criminal activities. Additionally, even other bosses rank better than *Ciro* and *Genny*, whose common feature is to be two young bosses.

Character Presence

We now examine the presence of characters. Since our analysis is based on the lexicon, we assess their presence by dialogues. In particular, we consider both the number of lines and the number of tokens. In Figure 10, *Genny* and *Ciro* dominate the scene, exhibiting figures roughly three times as large as their runner-ups under both indicators. Their heavy presence seems to

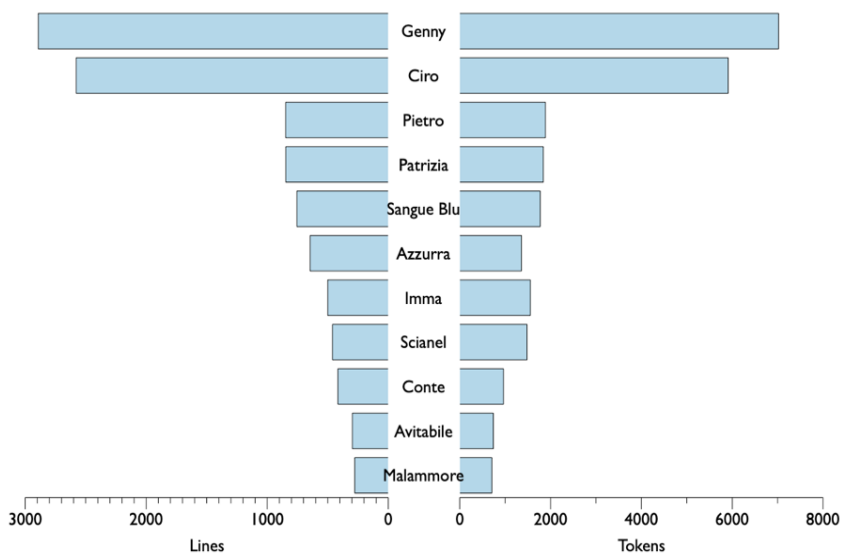


FIGURE 10
Characters' presence.

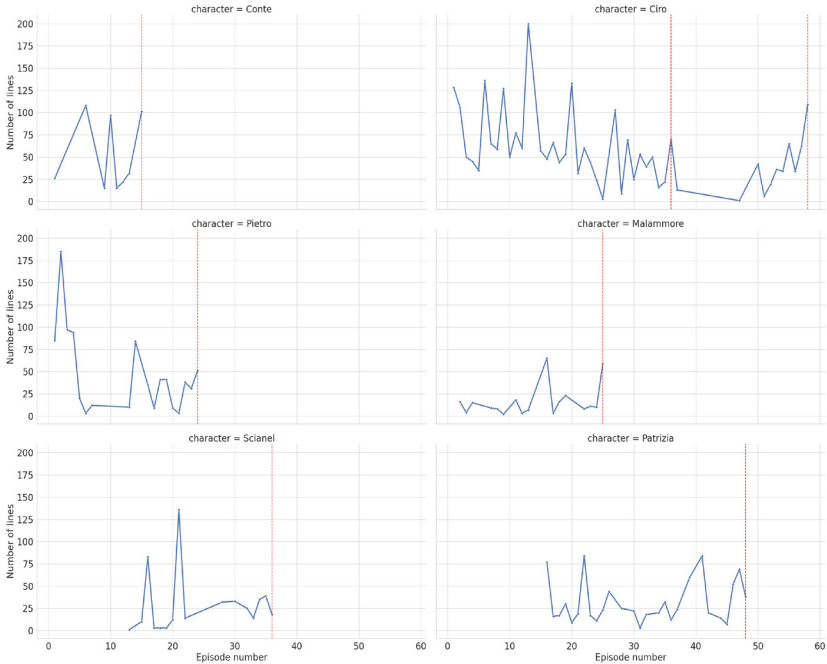


FIGURE 11
 Characters' presence across the series and the death factor.

show that though Don Pietro ranks higher in the criminal hierarchy for most of the series (being respectively Genny's father and Ciro's boss), the series is mainly centred on the two younger bosses.

We can also track the characters' presence throughout the series. In Figure 11, we see that some characters have an intermittent presence (see, e.g., both Conte and Scianel), and even the most prominent characters have a presence that is far from uniform (see the ups and down in Ciro's and Pietro's presence). On the same graph, we also notice what we might call "the death factor". For characters about to die, we observe a growing presence in the dialogues right before their death. In Figure 11, we have marked the episode where the character dies with a red label. We observe that death happens right at a local or global peak of presence. We observe a local peak for Pietro, Ciro (his first death before his apparent resurrection), Scianel, and Patrizia. Instead, we observe a global peak for Conte, Malammo'. It is

a narrative strategy that naturally creates a rising tension designed to capture the viewer's attention, a lexical hook, reminding us of the discursive hooks of narrative strategies in contemporary TV series scripts mentioned by Kristin Thompson (Thompson 2012).

Character Identification

Previously, we have seen that the lexica of some characters exhibit a degree of similarity. This observation significantly reduces the hope of identifying characters based on their lexicon. Here we investigate a character identification approach based on cosine similarity.

Cosine similarity is a measure of similarity between texts. We consider a space whose dimension is the overall number of unique words obtained after merging the two texts. Each text is treated as a term–frequency vector. The cosine similarity of the two texts is the cosine of the angle between the two vectors.

For character identification, we subdivide the overall text spoken by each character into two parts, which we call the train and test portions, with 80% and 20% proportions, respectively. We then examine the test portion of the vectors of character X, compute its cosine similarity with the training vector of all the characters, and attribute the test vector to the character with the highest cosine similarity. Character identification is correct if that test vector

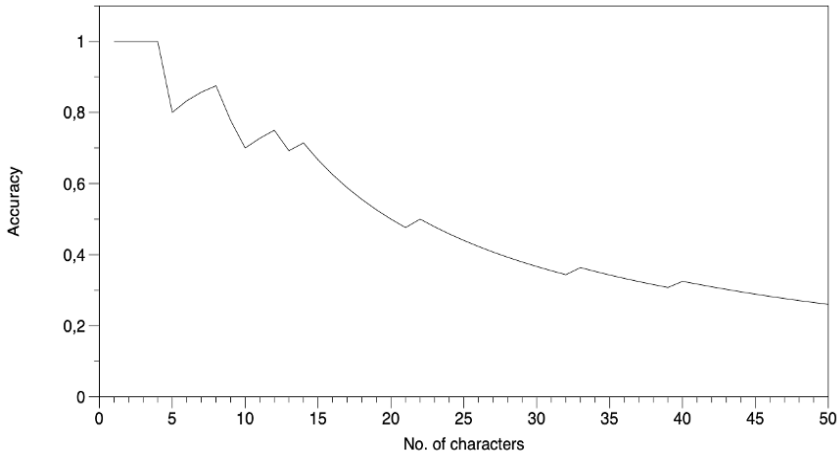


FIGURE 12
Characters' identification accuracy.

happens to be the test vector of character X. We evaluate the performance of this identification criterion by its accuracy, computed as the fraction of characters correctly identified. We expect the identification task to become more difficult as the number of potential characters to be identified grows. For that reason, we evaluate the algorithm's accuracy as we progressively increase the number of characters. We start with the most present characters (Genny) and then add one character at a time in order of decreasing presence (see Figure 10). When we have just two characters to identify, we may identify none, one, or two (i.e., with an accuracy of 0, 50%, or 100%).

The performance we get is shown in Figure 12. The accuracy is still larger than 80% when we have no more than eight characters and goes below 50% when we have more than 20 characters.

Conclusions

We started with a research question asking if we could identify a lexical signature of characters in *Gomorra*. A positive response would allow us to identify speakers based on their dialogues. We have shown early results in a research effort that is still in progress.

Our analysis of dialogues through simple similarity indices has shown that characters show both features common to the role of characters (e.g., the frequent use of certain words in bosses' dialogues) and features that can be exploited to identify individual characters (e.g., the type-to-token ratio to introduce the lexical variety of characters). Our first attempt shows that character identification is possible with a limited number of characters, and identification performances degrade as we increase the range of characters.

Our distant reading has, however, highlighted a lexical pattern that may be named bosses' code. Bosses appear to share a standard set of words that they frequently employ. That lexical corpus is typical of the gangster genre followed by the *Gomorra* series but also mimics the criminal reality that inspired *Gomorra* scriptwriters. The lexicon of Genny and Ciro is then both a real-life and a genre lingo, which we may dub Gomorrese.

Also, the distinguishing feature of complex tv series, which has been adopted as a reference to compare *Gomorra* with, is to narrate stories framed within a genre but set within a powerfully realistic context. The first series employing this mixture, at least within the gangster genre, was *The Sopranos*, created by David Chase and to which *Gomorra* has been

compared. *The Sopranos* ushered in the third golden age of American television and started the HBO model, followed by many producers-broadcasters (e.g., Sky): the story of the depressed gangster Tony Soprano moves within classic themes and stylistic elements of the American mafia film, but at the same time it also refers to the great American realist tradition.

Gomorra moves along the same lines. It is based on an investigative journalism book that is considered to belong to the new Italian epic. The term has been coined by the group of authors acting under the collective name of Wu Ming and refers to a trend in contemporary Italian fiction to deal with current Italian history by adopting the narrative conventions of a pop genre (Wu Ming 2009: 5-61).

Another distinguishing feature emerging from the analysis of dialogues is what we could call the death factor, i.e., the sudden increase of a character's relevance right before his/her death. *Gomorra*'s authors give dying characters the dead man walking stylistic feature: characters talk more when they are about to be silent forever.

On the other hand, the linguistic analysis confirms what the general audience may recognise: the Genny-Ciro pair's centrality. Though the story is rather coral, those characters represent two poles grabbing most of the video and dialogue presence. They also share the lexicon of older bosses. A distinguishing feature of that lexicon is that of being somewhat limited, using few words that often include the verbs *dovere* and *volere* (to want and to must). Those verbs embody the paradigm of the whole series: a perennial fight between obligations and wishes, which will remain unsolved for the two main characters.

This early quantitative analysis of dialogues in *Gomorra* has shown that we can uncover a set of patterns and models to build a layered interpretation of the series, following Moretti's suggestion to keep together the analysis of details and the drawing of a wider morphology.

BIBLIOGRAPHY

- Barra, Luca and Massimo Scaglioni (2015). "Saints, Cops and Camorristi. Editorial Policies and Production Models of Italian Tv Fiction." *SERIES – International Journal of TV Serial Narratives* 1(1): 65-76. <https://doi.org/10.6092/issn.2421-454X/5115>.
- Bednarek, Monika (2017). "The Role of Dialogue in Fiction." In *Pragmatics of Fiction* edited by Miriam H. Locher and Andreas Jucker, 129-158. Berlin-Boston: De Gruyter Mouton.
- Bednarek, Monika (2018). *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge: Cambridge University Press.
- Benvenuti, Giuliana (2017). *Il brand Gomorra. Dal romanzo alla serie TV*. Bologna: Il Mulino.
- Bernardelli, Andrea (2017). "Melodramma e realismo in Gomorra. La serie." *Ocula* 18: 1-14. <https://doi.org/10.12977/ocula85>.
- Beveridge, Andrew and Jie Shan (2016). "Network of Thrones." *Math Horizons* 23(4): 18-22. <https://doi.org/10.4169/mathhorizons.23.4.18>.
- Bolter, Jay David and Richard Grusin (1999). *Remediation: Understanding New Media*. Cambridge: MIT Press.
- Bost, Xavier, Vincent Labatut, Serigne Gueye and Georges Linarès (2016). "Narrative Smoothing: Dynamic Conversational Network for the Analysis of TV Series Plots." In *DyNo: 2nd International Workshop on Dynamics in Networks, in conjunction with the 2016 IEEE/ACM International Conference ASONAM*. <https://doi.org/10.1109/ASONAM.2016.7752379>.
- Chao, Dingding, Taro Kanno, Kazuo Furuta and Chen Lin (2018). "Representing Stories as Interdependent Dynamics of Character Activities and Plots: A Two-Mode Network Relational Event Model." *Digital Scholarship in the Humanities* 34(3): 471-81. <https://doi.org/10.1093/lcf/fqy062>.
- De Pascalis, Ilaria (2018). "An Italian Ecosystem: *Gomorra*." In *Reading Contemporary Serial Television Universes*, edited by Paola Brembilla and Ilaria De Pascalis, 113-27. New York-London: Routledge.
- Dusi, Nicola (2019). "Adapting, Translating and Reworking *Gomorra*." *Adaptation* 12(3): 222-39. <https://doi.org/10.1093/adaptation/apz007>.
- Edwards, Michelle, Lewis Mitchell, Jonathan Tukey and Matthew Roughan (2018). "The One Comparing Narrative Social Network Extraction Techniques." *arXiv*:1811.01467. <https://doi.org/10.1109/ASONAM49781.2020.9381346>.
- Erbaggio, Pierluigi (2015). "#*GomorraLaSerie*: Converging Audience and Enhanced Authors on Twenty-First-century Italian Screens." *Modern Italy* 20(4): 335-49. <https://doi.org/10.1080/13532944.2015.1096241>.

- Fronzetti Colladon, Andrea and Maurizio Naldi (2018). "Concentration Indices for Dialogue Dominance Phenomena in TV Series: The Case of *The Big Bang Theory*." In *International Conference on the Statistical Analysis of Textual Data*, 55-64. Cham: Springer.
- Fronzetti Colladon, Andrea and Maurizio Naldi (2019). "Predicting the Performance of TV Series through Textual and Network analysis: The Case of *Big Bang Theory*." *PLoS ONE* 14(11), e0225306. <https://doi.org/10.1371/journal.pone.0225306>.
- Fruitaldo, Antonio (2018). "(Re) Translating Culture-Bound Elements in *Gomorra-The Series*: A Corpus-Based Investigation into Relocated Identities." *Status Quaestionis* 15: 135-60. <https://doi.org/10.13133/2239-1983/14574>.
- Guerra, Michele, Sara Martin and Stefania Rimini (eds) (2017). "Un altro mondo in cambio. *Gomorra* fra teatro, cinema e televisione." *Arabeschi*, 9.
- Guerra, Michele, Sara Martin and Stefania Rimini (eds) (2018). *Universo Gomorra. Da libro a film, da film a serie*. Milano-Udine: Mimesis.
- Hammond, Michael and Lucy Mazdon (eds) (2005). *The Contemporary Television Series*. Edinburgh: Edinburgh University Press.
- Heimerl, Florian, Steffen Lohmann, Simon Lange and Thomas Ertl (2014). "Word Cloud Explorer: Text Analytics Based on Word Clouds." In *2014 47th Hawaii International Conference on System Sciences*, 1833-42. <https://doi.org/10.1109/HICSS.2014.231>.
- Holdaway, Dom (2014). "Osservazioni sulla retorica di *Gomorra*." In *Un nuovo cinema politico*, edited by Luciana Arcangeli, William Hope and Silvana Serra, 198-210. Leicester: Troubadour.
- Jayannavar, Prashant Arun, Apoorv Agarwal, Melody Ju and Owen Rambow, (2015). "Validating Literary Theories Using Automatic Social Network Extraction." In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 32-41.
- Jenkins, Harry (2006). *Convergence Culture: Where Old and New Media Collide*. New York-London: New York University Press.
- Labatut, Vincent and Xavier Bost (2019). "Extraction and Analysis of Fictional Character Networks: A Survey." *ACM Computing Surveys (CSUR)* 52(5): 1-40.
- Martin, Brett (2013) *Difficult men: Behind the scenes of a creative revolution: From The Sopranos and The Wire to Mad Men and Breaking Bad*. New York: Penguin.
- Martin, Sara (2017). "Il fascino sfacciato dei boss." In *Arabeschi* 9. *Un altro mondo in cambio. Gomorra fra teatro, cinema e televisione*, edited by Michele Guerra, Sara Martin and Stefania Rimini, 9: 238-41.
- McCabe, Janet and Kim Akass (eds) (2007). *Quality TV: Contemporary American Television and Beyond*. London: IB Tauris.
- Milkowska-Samul, Kamila (2017). "Per una nuova immagine delle donne in TV italiana. Alcune riflessioni in base alla fiction *Gomorra*." *International Journal for Linguistic, Literary and Cultural Research* 2(3): 193-203.
- Mittel, Jason (2015). *Complex tv: The Poetics of Contemporary Television Storytelling*. New York: New York University Press.
- Moretti, Franco (2005). *La letteratura vista da lontano*. Torino: Einaudi.
- Moretti, Franco (2022). *Falso movimento. La svolta quantitativa nello studio della letteratura*. Milano: Nottetempo.

- Naldi Maurizio and Paola Dalla Torre (2022). “Uncovering the Narrative Structure of *Breaking Bad* through a Multi-Dimensional Quantitative Analysis.” *SERIES – International Journal of TV Serial Narratives* 8(1): 17-32. <https://doi.org/10.6092/issn.2421-454X/14503>.
- Napoli, Antonella and Mario Tirino (2015). “*Gomorra* Remixed: Transmedia storytelling tra politiche di engagement, mainstream e produttività del fandom.” *SERIES – International Journal of TV Serial Narratives* 1(2): 191-204. <https://doi.org/10.6092/issn.2421-454X/5904>.
- Napoli, Antonella and Mario Tirino (2016). “Senza pensieri. *Gomorra – La serie*: dal contesto produttivo alle audience della Rete, fenomenologia di un processo culturale, transmediale.” *Mediascapes Journal* 7: 102-14.
- Pescatore, Guglielmo and Marta Rocchi (2020). “*Gomorra*: La serie, la formula del realismo.” *L'avventura. International Journal of Italian Film and Media Landscape* 6(Fascicolo Speciale): 87-108.
- Raffi, Francesca (2017). “*Gomorra – The Series* Flies to the UK: How Is *Gomorra*'s World Rendered in English Subtitles?” In *Transnational Subject. Linguistic Encounters. Vol. 2. Selected Papers from XXVII AIA Conference*, edited by Liliana Landolfi, Eleonora Federici and Flavia Cavaliere, 179-93. Napoli: Liguori.
- Renga, Dana (2015). “Making Men in *Gomorra: La serie*.” *L'avventura. International Journal of Italian Film and Media Landscape* 1: 105-20.
- Renga, Dana (ed.) (2016). “*Gomorra: La Serie*: Beyond Realism.” *The Italianist* 36(2): 287-354.
- Russo, Paolo (2017). “Storylining Engagement with Repulsive Antiheroes. Towards a Cognitive Poetics of TV Serial Drama Narrative: The Case of *Gomorra – The Series*.” *Journal of Screenwriting* 8 (1): 5-21.
- Scaglioni, Massimo and Luca Barra (eds) (2013). *Tutta un'altra fiction. La serialità pay in Italia e nel mondo*. Roma: Carocci.
- Thompson, Kristin (2012). *Storytelling. Forme del racconto tra cinema e televisione*. Soveria Manelli: Rubbettino.
- Vala, Hardik, Davide Jurgens, Andrew Piper and Derek Andrew (2015). “Mr. Bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the Difficulty of Detecting Characters in Literary Texts.” In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 769-74.
- Variano, Angelo (2019). “Il furbesco seriale della fiction: La lingua di *Gomorra: La Serie*.” *Lingue e Linguaggi* 30: 285-305. <https://doi.org/10.1285/122390359V30P285>.
- Waumans, Michael C., Thiabaut Nicodème and Hugues Bersini (2015). “Topology Analysis of Social Networks Extracted from Literature.” *PLoS ONE* 10(6), e0126470. <https://doi.org/10.1371/journal.pone.0126470>.
- Wu Ming (2009). *New Italian Epic. Letteratura, sguardo obliquo, ritorno al futuro*. Torino: Einaudi.

LOOKING FOR LEXICAL SIGNATURES IN GOMORRAH



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.a8ccd1c0>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

4. Women Behind the Scenes

A Data-Driven Approach to Analyse Female Labour
in Italian Television¹

Valentina Re and Marica Spalletta

◀ ABSTRACT

The aim of this work is analysing women's employment in key behind-the-scenes jobs in Italian TV crime dramas distributed from Fall 2015 to Summer 2022. This research has a two-fold objective: first, understanding female employment in the Italian television sector through the lens of crime dramas (*micro-level*); second, and more generally, setting an analysis model that can be replicated with respect to other genres, periods, countries or typology of datasets (*macro-level*). Since the sample analysis we conducted revealed data and trends consistent with analyses of larger datasets, we can argue that the original findings obtained may extend beyond the boundaries of the crime genre. Despite some positive trends, data about female employment in Italian television production show a more unbalanced scenario with stronger gender inequalities compared to both European averages for TV fiction and the Italian film sector.

KEYWORDS

Gender equality; female labour; TV series; Italian screen industry; data-driven approach.

¹ This study originates at the intersection of three research projects: the H2020 project DETECT – Detecting Transcultural Identity in European Popular Crime Narratives (2018-2021); the national project (PRIN 2020) The Atlas of Italian “Giallo”. Media History and Popular Culture (2022-2025); the CERV-2022-GE project GEMINI – Gender Equality through Media Investigation and New training Insights (2023-2025). This work is the fruit of genuine and intensive collaboration on all parts and aspects of the essay. Each of us was principally responsible for writing exactly half of the essay.

Introduction: Gender Equality, Data, Policies

As a fundamental EU value and a core EU objective, gender equality is surrounded by a strong policy background in all sectors of European societies – see for instance the *EU Gender Equality Strategy 2020-2025* and the *Council of Europe Gender Equality Strategy 2018-2023*. Given the influential role of media on how social and cultural norms relating to gender are perceived, specific measures have been progressively established to increase gender equality and diversity in the European audiovisual sector – e.g., the *Creative Europe MEDIA Programme (2021-2027)* and the *Eurimage – Gender Equality Strategy 2021-2023*.

The reference to “diversity” underlines the need to complement the gender equality perspective with an intersectional approach aimed at addressing discrimination on a systemic level, thus combining gender with other personal characteristics (e.g., race, religion, age, disability) which can pose a potential source of inequality, and promoting diversity as a value and a driver of growth and innovation in all areas of life. In this respect, several reports over the past five years have combined behind-the-scenes and on-screen perspectives to map gender issues in the audiovisual sector from the broader perspective of diversity and elaborate policy recommendations.² All these reports stress how the “availability of such knowledge is crucial for both creating awareness about existing gender issues as well as developing appropriate policies and good practices” while pointing out the “lack of systematic data” (EENCA 2020: 8). As clearly stated in the conclusions of

² See at least EAO 2021, EENCA 2020, Voices of Culture 2020, ERGA SG 4 2019, and EAO 2018. Concerning the impact on policy strategies and funding, see the BFI diversity standards as a paradigmatic case: <https://www.bfi.org.uk/inclusion-film-industry/bfi-diversity-standards> (last accessed 14-01-23).

the 2020 report *Achieving Gender Equality and Promoting Diversity in the European Audiovisual Sector, Good Practice Handbook*: “The high value of the regular production of statistics and qualitative analyses of the levels of representation of women and other under-represented groups on screens and in the audiovisual workforce has been demonstrated” (Tepper 2020: 69).

The first report released by the Italian Observatory on Gender Equality, set up by the Ministry of Culture in November 2021, goes precisely in the same direction. At the beginning of the last chapter, entitled “Uno sguardo ai dati” (“A look at data”), we can read:

The availability of data to monitor gender equality in the cultural sectors is still scarce and above all heterogeneous across sectors. Nevertheless, the available data, to which we will refer in the following pages [...], confirm strong gender inequalities and the need to promote a systematic collection of information that can shed light on the inequalities that have emerged, which appear difficult to unravel and which require targeted policies for the promotion of equality in all sectors of culture (MiC – Osservatorio sulla parità di genere 2022: 47; our translation).

The data to which the Italian Observatory refers concern exclusively female employment in the film sector, and more specifically come from: *Women Directors and Film Production. Addressing the Gender Gap and the Role of Festivals* (WIFTMI 2019), *Tutti i numeri del Cinema e dell'Audiovisivo* (MiC 2020), *Gender Balance in Italian Film Crews. Data and Research Policies* (MiC and Università Cattolica del Sacro Cuore 2022). Despite the huge popularity of television narratives and the “content arms race” in the age of the so-called “peak TV”, the only data currently available on gender equality in Italian television production are those released by the European Audiovisual Observatory (EAO) (Fontaine 2022), APA (Associazione Produttori Audiovisivi) and Fondazione Symbola (2022), and the public broadcaster Rai – the latter however limited to the on-screen perspective (Rai 2022).

We can similarly see a need for more data available in academic research. Although film and television studies are becoming increasingly concerned with data-driven approaches (Bielby 2014, Krijnen and Van Bauwel 2015, Liddy 2020), it is still difficult to find research based on a mixed methods approach (Creswell and Plano Clark 2011), where qualitative and quantitative approaches blend to offer qualitative analysis of quantitative data and vice versa. At the same time, these studies often tend to employ methodological triangulation models “within-method” or “between/across-method”,

where instead only the full integration of convergent and divergent perspectives would ensure the accuracy of researchers' assessments (Amaturo and Punziano 2016). Finally, it cannot be ignored that quantitative analyses provided by both academic research and public reports are often based on different datasets and frequently recur to different types of measuring, with the result that in most cases the research findings are not directly comparable because they are uneven both on the side of data collection and data analysis.

Aims & Methodology

It is precisely to fill these gaps that this study – which is part of a broader research aimed at analysing gender equality and diversity in Italian television production³ – combines data analysis and production studies both to analyse the gender equality issue and female labour in contemporary Italian TV crime drama (*micro-level*) and to set an analysis model that can be replicated, for example, with respect to other genres, periods or countries (*macro-level*).

While considering the extensive Italian and international literature on gender and media in both social sciences and television studies, this article is mainly concerned with methodologies and data evidenced by reports from stakeholders such as public broadcasters, streaming services, public service organisations and non-profit research organisations.

For this purpose, this study focuses on Italian TV crime dramas distributed from Fall 2015 to Summer 2022 (seven television seasons) as a paradigmatic case study.

The focus on the crime genre depends on well-established findings emerging from the academic literature of recent years. On the one hand, crime narratives provide the opportunity to analyse mainstream productions and popular narratives; as argued in *Transnational European Television Drama*, “crime drama [...] is clearly the most popular genre across Europe” (Bondebjerg et al. 2017: 223) and it “serves as a lens through which to observe the local, national and even transnational issues that are prevalent in

³ For a more comprehensive account of this ongoing research, which combines behind-the-scenes and on-screen perspectives, see our article “Unsuitable Jobs for Women: Women’s Behind-the-Scenes Employment and Female On-Screen Representation in Italian TV Crime Drama” (*Comunicazioni sociali*, 1, 2023, pp. 82-97) and the related report available at: <https://atlantedelgiallo.unilink.it/index.php/2022/11/18/professionii-inadatte-a-una-donna-report/> (last accessed 14-01-23).

a society” (Hansen et al. 2018: 1). On the other, the crime genre has been investigated by a rich tradition of feminist cultural and media studies that have explored the role of women as creators, characters, and consumers of crime narratives. In this respect, the crime genre has represented a privileged perspective to observe changes in gender roles, both in society and fictional worlds, and particularly the relationships between the representation of female investigators and changes in female condition in the broader socio-cultural context, especially in the workplace and domestic spaces.⁴

The time frame 2015-2022 is motivated both internationally and nationally. On the one hand, it allows comparing the evidence of this research with data provided by EAO about the presence of women among six audiovisual professional categories in TV fiction with a first broadcast or release in the EU27 plus UK between 2015 and 2020 (Fontaine 2022).⁵ On the other hand, at the national level, 2015 stands out as a turning point in the innovation processes implemented by Rai, the Italian PSB and leading provider of TV series, to respond to original productions by linear and non-linear pay TV players (Sky, Netflix, Amazon) and to the increasing circulation of production and narrative models from other European countries, and especially Nordic ones. More particularly, the TV series *Non uccidere* (*Thou Shalt Not Kill*, 2015-2018), aired on Rai 3 in September 2015, pushed Italian production into an international context thanks to its innovative visual style and its female protagonist, Valeria Ferro, a woman in her thirties and inspector of the homicide squad.

From a methodological point of view, the background analysis of existing reports has shown that they often tend to be based on vast amounts of data resulting from large datasets.⁶ Consistently with the macro objective of our research (namely, to set an analysis model that can then be replicated), we chose to adopt a sample methodology, which allows us to make infer-

⁴ See in particular: Dresner 2007, Gates 2011, Hoffman 2016, Klein 1995, Reddy 2003, and Turnbull 2014. For a literature review focused on the Italian context, see D’Amelio and Re 2021a.

⁵ See also Jiménez Pumares 2021. Pioneering research on gender equality in the Italian audiovisual sector, with particular reference to the film industry, has been conducted by the research project DEA from 2016 to 2019, see DEA 2016 and 2018. For more recent approaches focused on Italian cinema, see Fanchi and Tarantino 2021 and the research project CineAF: Women’s Films in Italy (1965-2015) (<http://cineaf.eu/>, last accessed 14-01-23).

⁶ For instance, the sample of the 2022 EAO report is composed of 37,138 TV films and episodes for directors (90% of all TV films and episodes produced) and 34,706 for writers (84% of the entire production).

ences about a selected sample corpus, expecting to make them generalisable (Corbetta 2015). In particular, the sampling procedure took into consideration the different methodological choices adopted by other reports working on a sample basis, merging them into an original model within which:

1. In accordance with the report *Boxed In 2019-2020* (Lauzen 2020), we analysed only one episode for each TV series distributed in the selected time span;
2. Following the report *Women Over 50* (Geena Davis Institute on Gender in Media 2021), we did not consider one randomly selected episode, but always the second episode of season 1 (01x02). This choice allows focusing on the first season of each series (which stands for the ‘original crew’) while considering how season premieres or pilot episodes, as well as season finales, may differ from regular season norms.

Behind-the-scenes jobs we tracked correspond to four main categories:

- *Above-the-line roles*: director(s), writer(s), delegate producer(s);⁷
- *Below-the-line roles*: cinematography, music, sound, editing, costume design, production design, special/visual effects, casting, make-up;
- Main producer(s);
- Producer(s) for the commissioner.

Data were collected in two steps. First, we selected all the Italian TV series (distributed in the predetermined time frame) labelled as ‘crime’ and ‘thriller’ (or in Italian *giallo* and *poliziesco*) in the IMDb database or, even otherwise labelled (e.g., ‘action’ or ‘drama’), referable to ‘crime fiction’ as a pragmatic set of features and expectations shared by producers and viewers and displayed in promotional strategies. At the end of this step, the selected

⁷ The high heterogeneity in crediting the various producer figures, which describe equally diverse production practices, made it necessary to make some methodological choices. As we wanted to focus on figures with decision-making and supervising roles, we distinguished between the main producer(s) (that is, the main financial supporter[s]), his/her main delegate(s) (in Italian *produttore delegato* or *produttore esecutivo*) and the producers representing the commissioner (e.g., the channel/platform of the first release, which is also the major funder). At this research stage, the tracking of behind-the-scenes jobs through the credits did not include the main cast because the actors’ role was addressed separately from the on-screen perspective and the analysis of characters. For more details see the report *Unsuitable Jobs for Women*, <https://atlantedelgiallo.unilink.it/index.php/2022/11/18/professionisti-inadatte-a-una-donna-report/> (last accessed 14-01-2023).

Italian TV crime dramas amounted to 62. The second step consisted of collecting data concerning the selected key roles. To this purpose, we considered the opening credits of each episode, integrating them, when needed (e.g., visual and special effects supervisor, producer), with final credits or the information provided by the IMDb. Considering all the names in the credits, the analysed corpus amounts respectively to 1,158 mentions referring to the sex each name conveys (that is, counting each name every time the credits quoted it) and to 759 professionals with respect to the individual they refer to (that is, counting each name only one time).

The collected data were then processed and subjected to a double type of analysis, monovariate – based on the study of each variable individually, through descriptive statistical analysis tools – and bivariate – more properly exploratory in nature, aimed at verifying the existence of any association between the variables, and to reach a proper interpretation of the same (Marradi 2007).

Patterning Italian Crime Dramas

The analysis of women's employment in key behind-the-scenes roles in Italian crime dramas requires to preliminarily patterning the object of analysis (Tab. 1). To this purpose, we have considered five main elements:

1. *First release*: when (before/since Fall 2015) and on which channel/platform (Rai, the commercial broadcaster Mediaset, the Pay TV Sky, Netflix, Amazon Prime Video);
2. *Production and format*: if it is an international co-production, an original production, a literary adaptation (from a female or male novelist), or a remake; episode duration (less/more than 65');⁸
3. *Setting*: the period (present or past – before the 2000s) and the main location where the story unfolds;
4. *Narrative form*: episodic or serial narrative;⁹
5. *Protagonist(s)*: male-based, female-based, duo-based or team-based (ensemble).

⁸ For this distinction see Fontaine 2022.

⁹ The binary distinction between long-term arc storytelling and stand-alone episodes refers to the form prevailing in every TV series and does not strictly categorise it. In fact, in current practice, the two forms are increasingly combined, and especially the “mystery of the week” traditional model is now commonly associated with a more or less complex and relevant horizontal storyline.

TV SERIES	YEARS	SEASONS	EPISODES	CHANNEL/PLATFORM
<i>Baby</i>	2018-2020	3	18	Netflix
<i>Bang bang baby</i>	2022-	1	10	Amazon Prime Video
<i>Bella da morire</i>	2020	1	8	Rai 1
<i>Blanca</i>	2021-	1	6	Rai 1
<i>Blocco 181</i>	2022-	1	8	Sky Atlantic
<i>Catturandi – Nel nome del padre</i>	2016-	1	12	Rai 1
<i>Diavoli</i>	2020	2	18	Sky Atlantic
<i>Don Matteo</i>	2000-	13	265	Rai 1
<i>Fosca Innocenti</i>	2022-	1	4	Canale 5
<i>Fuoco amico TF45 – Eroee per amore</i>	2016	1	8	Canale 5
<i>Giustizia per tutti</i>	2022	1	6	Canale 5
<i>Gli orologi del diavolo</i>	2020	1	8	Rai 1
<i>Gomorra: La serie</i>	2014-2021	5	58	Sky Atlantic/Sky Cinema
<i>I bastardi di Pizzofalcone</i>	2017-	3	18	Rai 1
<i>I delitti del BarLume</i>	2013-	9	18	Sky Cinema
<i>I misteri di Laura</i>	2015	1	8	Canale 5
<i>Il bosco</i>	2015	1	4	Canale 5
<i>Il cacciatore</i>	2018-2021	3	28	Rai 2
<i>Il capitano Maria</i>	2018	1	4	Rai 1
<i>Il commissario Montalbano</i>	1999-2021	15	37	Rai 2/Rai 1
<i>Il commissario Ricciardi</i>	2021-	1	6	Rai 1
<i>Il giovane Montalbano</i>	2012-2015	2	12	Rai 1
<i>Il miracolo</i>	2018	1	8	Sky Atlantic
<i>Il processo</i>	2019	1	8	Canale 5
<i>Il re</i>	2022-	1	8	Sky Atlantic
<i>Il silenzio dell'acqua</i>	2019-2020	2	8	Canale 5
<i>Il sistema</i>	2016	1	6	Rai 1
<i>Imma Tataranni - Sostituto procuratore</i>	2017-	2	14	Rai 1
<i>Io ti cercherò</i>	2020	1	8	Rai 1
<i>L'allieva</i>	2016-2020	3	35	Rai 1
<i>L'ispettore Coliandro</i>	2006-	8	34	Rai 2
<i>L'alligatore</i>	2020	1	8	Rai 2
<i>Lonore e il rispetto</i>	2006-2017	5	32	Canale 5
<i>La fuggitiva</i>	2021	1	8	Rai 1
<i>La narcotici</i>	2011-2015	2	12	Rai 1
<i>La porta rossa</i>	2017-	2	24	Rai 2
<i>La strada di casa</i>	2017-2019	2	24	Rai 1
<i>Le indagini di Lolita Lobosco</i>	2021-	1	4	Rai 1
<i>Màkari</i>	2021-	2	7	Rai 1
<i>Maltese – Il romanzo del commissario</i>	2017	1	4	Rai 1
<i>Masantonio - Sezione scomparsi</i>	2021-	1	10	Canale 5
<i>Monterossi - La serie</i>	2022	1	6	Amazon Prime Video
<i>Nero a metà</i>	2018-	3	36	Rai 1
<i>Non mentire</i>	2019	1	3	Canale 5
<i>Non mi lasciare</i>	2022	1	8	Rai 1
<i>Non uccidere</i>	2015-2018	2	36	Rai 3/Rai 2/Rai Premium
<i>Petra</i>	2020-	1	4	Sky Atlantic/Sky Cinema
<i>Pezzi Unici</i>	2019	1	12	Rai 1
<i>Provaci ancora prof!</i>	2005-2017	7	46	Rai 1
<i>Rocco Schiavone</i>	2016-	4	16	Rai 2
<i>Romanzo Siciliano</i>	2016	1	8	Canale 5
<i>Rosy Abate – La serie</i>	2017-2019	2	10	Canale 5
<i>Solo</i>	2016-2018	2	8	Canale 5
<i>Solo per amore</i>	2015	2	18	Canale 5
<i>Sorelle</i>	2017	1	6	Rai 1
<i>Sotto copertura</i>	2015-2017	2	10	Rai 1
<i>Squadra antimafia – Palermo oggi</i>	2009-2016	8	74	Canale 5
<i>Squadra mobile</i>	2015-2017	2	32	Canale 5
<i>Suburra - La serie</i>	2017-2019	3	24	Netflix
<i>Un passo dal cielo</i>	2011-	6	80	Rai 1
<i>Vite in fuga</i>	2020	1	12	Rai 1
<i>Vostro onore</i>	2022	1	8	Rai 1

TABLE I

The analysed corpus.

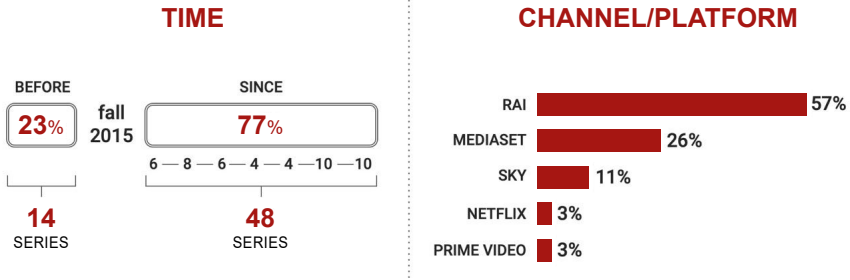


FIGURE 1
First release.

Crime dramas distributed since Fall 2015 consist of 48 TV series (77%), while 14 cases (23%) started before Fall 2015 and were renewed in the 2015-2016 season or the following ones. Figure 1 shows that over the last two seasons the number of new productions has significantly increased, thus proving the popularity of the crime genre. Concerning channels and platforms, about half of the corpus (57%) consists of crime dramas co-produced by Rai, while 1 in 4 (26%) comes from the commercial broadcaster Mediaset. The remaining ones are co-produced by Sky (11%), Netflix (3%), and Amazon Prime Video (3%). Comparing the performance of the broadcasters in the shift before/since Fall 2015, an inverse trend emerges: the number of Rai TV series grows (from 50% to 59%) while Mediaset productions decline (from 36% to 23%) (Fig. 1).

Moving to the second feature (*production and format*), only two crime dramas are international co-productions, *Diavoli* (Italy/France/UK, 2020-) and *Gli orologi del diavolo* (Italy/Spain, 2020). 3 TV series in 5 (60%) represent original productions, while 35% consist of adaptations (male novelists prevail over female ones – 71% vs 29%); 3 cases (*I misteri di Laura*, 2015, *Non mentire*, 2019, and *Vostro onore*, 2022) are remakes of original crime dramas produced in Spain (*Los misterios de Laura*, 2009-2014), UK/USA (*Liar*, 2017-2020) and Israel (*Kvodo, Your Honor*, 2017). A significant transformation concerns the episode duration: overall, crime dramas with episodes longer than 65' are more than half (52% vs 48%). However, splitting the data shows an inverse trend: series released before Fall 2015 present longer episodes (71% vs 29%), while more recent series prefer shorter (54% vs 46%). Given the widespread popularity of this format abroad, we can interpret the increasing use of the less than 65' format in terms of a growing internationalisation process of Italian TV series (Fig. 2). Again, *Non ucci-*

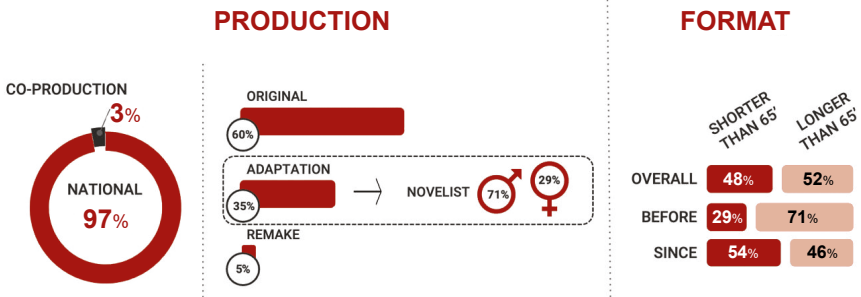


FIGURE 2
Production and format.

dere can be interpreted as a paradigmatic case. When renewed for a second season, episodes are doubled (from 12 to 24), but their length is reduced to 50 minutes at the request of foreign buyers.¹⁰ In fact, despite not achieving good ratings in Italy, the show circulated quite well abroad – in France (*Squadra criminale*) and Germany (*Die Toten von Turin*) on Arte and on the British Channel 4' Walter Presents streaming service (also available in the US from 2017) for UK and US audiences.

Concerning the third feature (*setting*), almost all the analysed crime dramas (90%) are set in the present, while 10% look at the past both for narrative choices (i.e., *Il commissario Ricciardi*, 2021-, and *Bang Bang Baby*, 2022-) and when the plot refers to actual events dated in specific periods of Italian history (e.g., the “Italian Mafia” which recurs in *Il cacciatore*, 2018-2021, and *Maltese – Il romanzo del commissario*, 2017). Regarding locations, Rome peaks as the “Capital” of Italian crime dramas (13 productions), followed by Sicily (where Montalbano’s fictional Vigata is located); the only TV series set abroad is *Diavoli*, with a plot unfolding primarily in London. Crime dramas released since 2015, however, show the development of more diversified location strategies, with new productions set in Basilicata, Calabria, Liguria, Lombardia, Piemonte, Puglia, Valle d’Aosta, and Veneto (Fig. 3).

Concerning the *narrative form*, the serial model prevails over the episodic one (60% vs 40%). However, data show an interesting shift in the time frame: before 2015, in fact, 57% of crime dramas belonged to the episodic model, while starting from 2015, the serial form prevailed (65% vs 35%)

¹⁰ See the interview with creator Sergio Corbucci available on Rai streaming portal RaiPlay, <https://www.raipplay.it/> (last accessed 14-01-23).

TIME

PRESENT



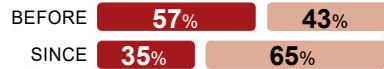
LOCATION



OVERALL



BEFORE | SINCE 2015



FIGURES 3, 4
Setting and Narrative form.

(Fig. 4). Moreover, as already pointed out, the episodic case is increasingly associated with a more or less complex and relevant horizontal storyline.

Finally, an interesting transformation also emerges regarding the *protagonist's role*. Overall, male protagonists peak with 40%, followed by female protagonists (31%), team-based (ensemble) narratives (21%), and duo-based narratives (8%).¹¹ However, comparing TV series released before and since 2015, ensemble narratives show a descending curve (36% vs 17%) to the advantage of female protagonists, which grew from 21% to 33%. The “decline” of the team-based model, which has mainly char-

¹¹ In all the analysed series, every duo consists of a man and a woman. The only case in which the duo involves two characters of the same sex/gender (male/man) refers to the Fortebracci brothers in *L'onore e il rispetto* (2006-2017).

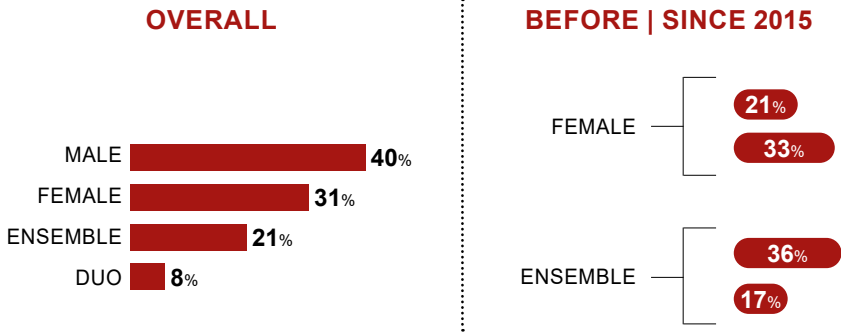


FIGURE 5
Protagonist(s).

acterised Mediaset productions in the past, can be partially explained by the decreasing number of Mediaset productions but also by the decreasing relevance as strategic drivers of internationally successful Sky productions such as *Romanzo criminale – La serie* (2008-2010) and *Gomorra – La serie* (2014-2021). On the other hand, the growth of female protagonists and the focus on female narratives can be linked to Rai’s innovative production strategies mentioned above. Focusing on the two broadcasters, there is no difference between Rai and Mediaset concerning the share of crime dramas with female protagonists (32% vs 31%), and the same applies to ensemble narratives (14% vs 13%). However, while in Rai TV series stories with male protagonists prevail (51% vs 31%), the share of Mediaset series featuring a duo stands out to be much higher (25% vs 3%). Indeed, whereas the duos are usually composed of males and females, there would seem to be more room for women in Mediaset series than in Rai (Fig. 5).¹²

Female Employment in Behind-the-Scenes Roles

Before introducing the research’s main findings, a methodological premise is due. In fact, referring to the credits does not allow to distinguish between sex and gender identity. Therefore, the distinction between female/women

¹² For an overview of Italian TV policewomen and the role of female characters in the team-based model see Buonanno 2012 (especially chapter 6, “Le ragazze con la pistola”). For an overview of recent productions see D’Amelio and Re 2021b.

professionals and male/man professionals was made based on the credited personal name and the sex with which this name is commonly associated in Italian culture.¹³ Of course, this choice in no way entails a binary approach to gender issues. Only further qualitative research, based on surveys and interviews with professionals, could make it possible to investigate, also from the behind-the-scenes perspective, the relations between sex, gender identity, and sexual orientation and how these aspects impact the working conditions in the audiovisual sector.

Measuring Female Employment by Mentions

To understand women's employment in key behind-the-scenes roles, we refer first to the mentions' parameter. Therefore, we analysed the overall 1,158 credits, measuring the female presence with respect to both the total tracked credited roles (*how many women are overall involved in the crews in comparison with men?*) and all the analysed crime dramas (*how many TV series include women in their crews?*).

General data show predominantly male crews: in fact, the overall female presence represents 31% of the 1,158 analysed credited roles, while the male presence amounts to 69%; therefore, for each woman involved, there are at least two men. On the contrary, only one series (*Catturandi – Nel nome del padre*, 2016) does not include any woman in its crew (Fig. 6).

Male employment prevails in the above-the-line roles, with no relevant differences between series released before or since 2015. In the analysed time frame, in fact, female directors represent 6%, and only three women (Cinzia Th Torrini, Maria Sole Tognazzi and Rossella Izzo) are directors of 4 series; finally, three series in 4 also present a female protagonist when there is a female director. Compared to direction, female employment in scriptwriting rises to 25%, but male writers still prevail (75%). This trend is confirmed by looking at the series: although in 32 of 62 crime dramas there is a female writer, only two series (*La narcotici*, 2011-2015, and *Vostro*

¹³ While in the international scenario "Andrea" is both a male and female name, in Italy it is mainly used to identify male persons; therefore, all crew members with that name have been coded as male. On the other hand, for personal names not belonging to the Italian tradition (e.g., Emita), pseudonyms (e.g., Salmo) and bands (e.g., Calibro 35), further research was carried out to identify their sex. Finally, based on the biographical names given in the credits, it was impossible to identify the share of intersex, which would only be identifiable through surveys with direct answers.



FIGURE 6
Overall female employment.

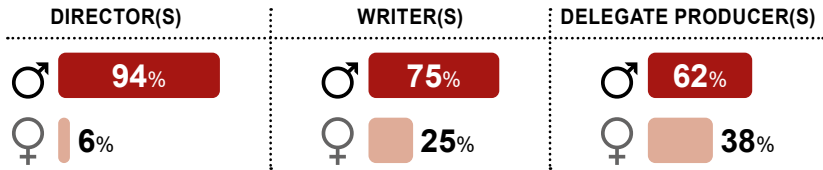


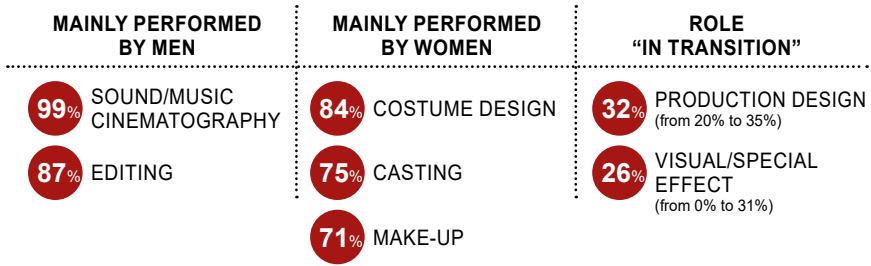
FIGURE 7
Female employment in above-the-line roles.

onore) are entirely written by women and, more in general, a male majority characterises writing teams. The female presence in the above-the-line roles peaks in the role of delegate producer: although men still prevail, women account for 38%. This larger female employment is confirmed by looking at the second parameter: women work as delegate producers in 20 of 43 series in which this role is conferred (Fig. 7).

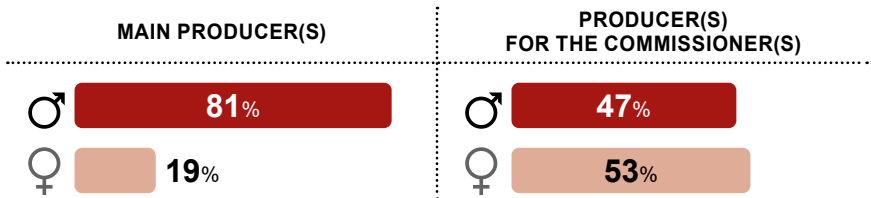
Concerning below-the-line roles, data analysis shows three different trends:

1. *Roles mainly performed by men*: cinematography (99%), music (99%), sound (99%) and editing (87%). Only one series credits women in cinematography (*Il miracolo*, 2018), sound (*Gomorra*) and music (*Il processo*, 2019), while editing boasts a more significant female presence: 9 of 62 series credit at least one woman in their editing crews;
2. *Roles mainly performed by women*: costume design (84%), casting (75%) and make-up (71%). Respectively, 53 of 62 and 44 of 59 series involve at least one woman in costume design and make-up, while 50 of 62 series credit a female presence in casting roles;
3. *Roles 'in transition'*: production design (32%) and visual/special effects (26%). These roles show an increase in female employment from crime dramas released before or since Fall 2015: female production designers shifted from 20% to 35%, while female visual/special effects supervisors grew from 0% to 31%. Women serve as production

WOMEN BEHIND THE SCENES



FIGURES 8
Female employment in below-the-line roles.



FIGURES 9
Female employment in production roles.

designers in 21 of 62 series, while supervising visual/special effects is a matter for women in 15 of 52 crime dramas (Fig. 8).

Regarding production roles, women have more opportunities as producers for the commissioner than as main producers (53% vs 19%). However, looking at time distribution, both roles show a negative trend: since 2015, the share of female main producers has dropped from 32% to 15%, and commissioners have reduced the number of their female producers from 71% to 48%. Nevertheless, 21 of 58 series present at least one woman in their production team, while series in which women are employed as producers for the commissioner amount to 34 of 44 (Fig. 9).

Measuring Female Employment by Individuals

In the second step of the research, we referred to a corpus consisting of 759 professionals, obtained by considering each mention by the individual it refers to.

Overall, data analysis confirms that, for each woman, there are at least two men involved in the crews: in fact, the female presence stands at 32%

of the 759 professionals, in line with the 31% of the 1,158 credited roles. At the same time, the individuals' perspective shows two main trends: on the one hand, female professionals involved in more than one production team are less than men, with the only exception of those roles we have defined as "mainly performed by women", where the repetitiveness in the role is also "a matter of women"; on the other hand, the employment in more than one crew negatively affects the percentage of women even in those few roles apparently "less unsuitable for women".

Referring to above-the-line roles, the number of professionals tends to decrease because they are often involved in 2 or more series as directors, writers or delegate producers; on the contrary, the share of women employed in the crews tends to be confirmed. Referring to directors and delegate producers, the overall number decreases respectively from 63 to 47 and from 85 to 66 (because 12 professionals and 14 delegate producers are involved in two or more series), but the man/woman percentage does not change (94% *vs* 6% and 62% *vs* 38%). Concerning writers, they amount to 122 units instead of 158 (23 professionals are involved in two or more writing teams and, among them, there are five women), but women slightly increase from 25% to 27%.

Looking at below-the-line roles, men prevail in photography, music and sound, which are characterised by a significant recurrence of the same professionals: directors of photography working in two or more productions are 13 out of 49, composers are 15 out of 43 (and among them Andrea Farri, Ralf Hildenbeutel and Savio Riccardi recur in more than five series), and sound directors are 16 out of 46. As for editing, the role appears always mainly performed by men, albeit to a lesser extent than in the previous three roles. In fact, the total number of editors decreases from 78 to 46 because 15 professionals work in more than one series, and of these two professionals (Alessandro Heffler and Pietro Morana) are engaged in 6 and 7 series respectively; however, only two women appear to be engaged in more than one production team, compared to 13 men, which explains why, in percentage terms, the female presence increases from 13% to 17%.

Looking at the roles mainly performed by women, the number of make-up artists drops from 67 to 49 (due to 13 professionals engaged in more productions), while the share of female employment is confirmed (71%). Percentages significantly decrease with reference to the roles of costume designer and casting director (from 84% to 80% and from 75% to 66%, respectively). However, this reduction is directly proportional to the greater "presence" of women in more productions: 13 out of 49 costume designers,

in fact, work in more than one series, and in 11 cases they are women (with Marina Roberti appearing in four productions); the same goes for casting, where 18 out of 44 directors are engaged in more than one series, and of these 15 are women. The high number of female professionals involved in casting in multiple series (as in the cases of Elisabetta Curcio and Adriana Sabbatini, engaged respectively in five and six productions) means that, percentage-wise, female employment drops from 75% to 66%. Finally, “roles in transition” show two different trends. In fact, the number of production designers drops from 69 to 48 due to 15 professionals working in more than one series (in particular, Valerio Girasole engaged in 4 different productions); however, the fact that men working in multiple series are more than women (11 vs 4 women) causes female employment to increase from 32% to 38%. On the contrary, the share of women engaged as special/visual effects supervisors drops from 26% to 20%, and this is because women engaged in more production are as many as men (among them, Monica Galantucci is the record woman of attendance with seven series to her credit).

Finally, with reference to production roles, the number of main producers decreases significantly from 129 to 65, and this is due to 26 professionals engaged in the production of more than one series (the record man is Carlo Degli Esposti, with seven productions to his credit); among these, however, only five are women, and this determines an increase in the percentage of female employment (from 19% to 25%). An inverse trend characterises the role of the producer for the commissioner: in fact, the number of professionals stands at 40 (compared to 83 in the previous measurement), 21 of whom are engaged in multiple productions; about half of these professionals (precisely 12) are women, resulting in a minimal change in the level of female employment, which stands at 58%.

Conclusions

This research has a two-fold objective: first, understanding female employment in the Italian television sector through the lens of crime dramas (*micro-level*); second, and more generally, setting an analysis model that can be replicated with respect to other genres, periods, countries or typology of datasets (*macro-level*).

Starting from the first goal, the data emerging from this study about the discrete amount of women in the role of scriptwriters (25%), as well as the extremely poor presence of women in the roles of director (6%), cinematog-

rapher (1%) and composer (1%), are in line with those on Italian TV fiction provided in the EAO 2022 report – female directors between 10 and 15%, writers between 30 and 35%; cinematographers and composers less than 5%.¹⁴ These data show a more unbalanced scenario with stronger gender inequalities compared to European averages (17% directors, 36% writers, 7% cinematographers, 5% composers) and data for the US context provided in the report *Boxed In 2019/20* (referring to both scripted and non-scripted shows): 30% directors, 36% writers, 8% cinematographers.

Since the sample analysis we conducted revealed data and trends consistent with analyses of larger datasets, we can argue that the original findings obtained may extend beyond the boundaries of the crime genre and may be compared with data about the Italian film sector.

In this respect, the increase of female production designers (from 20% in TV series started before Fall 2015 to 35% in series distributed since Fall 2015) and VFX/special effects supervisors (from 0% to 31%) shows a positive trend. However, data on the different production roles reveal a bleaker scenario, with strong gender inequalities in top positions. Women represent only 19% of main producers and tend to occupy more functional, albeit relevant, roles such as producer for the commissioner (53%) and *produttore delegato* (38%) – this latter figure is in line with the general figure for Italian female producers provided by EAO, between 30% and 35% (European average 44%).

Comparison with the Italian film sector reveals a similar picture, with the same preponderance of female professionals in roles traditionally associated with women (such as make-up) and significant disparities in other key roles, although far less pronounced than in the television sector – especially for composers, cinematographers and editors. Over the five years from 2017 to 2021, the share of female film directors rose from 14% to 19%, the share of female screenwriters from 22% to 23%, the share of female composers from 12% to 16%, the share of female cinematographers from 8% to 10%, while the share of female editors remained stable (27%). The figure for women film producers rose from 25 to 26%, but without reference to the more specific role played in the production team (MiC and Università Cattolica del Sacro Cuore 2022; MiC – Osservatorio sulla parità di genere 2022: 49).

¹⁴ Note that in all these cases the share of female professionals is weighted by the number of professionals in that role of each episode.

Compared to the first step, the second step of the research (measuring female employment by individuals) confirms the general average of women's employment. At the same time, a significant "qualitative" gender difference stands out: if men tend to be engaged more frequently in two or more productions (especially in below-the-line roles), "repetitiveness" does not seem to be "a matter for women", except for those roles "mainly performed by women". However, paradoxically, where this "repetitiveness" occurs, the share of female employment tends to fall compared to the male one.

In conclusion, we can argue that further analyses of gender equality in the Italian television sector, focusing on other genres and formats and with a longer time span, are strongly recommended, both to get a more comprehensive picture of Italian television production and to allow a more in-depth and systematic comparison with the film sector and other European and US production cultures. More particularly, the use of larger datasets could also allow an analysis of how gender issues and discriminations relate to creative clusters and professional collaboration networks in Italian television production – following the models of application of social network analysis to Italian cinema offered by Holdaway (2017) and Fanchi and Tarantino (2021).

At the same time, it is necessary to conduct a thorough methodological investigation that considers at least two main aspects. Firstly, it is essential to assess which data collection and analysis protocols are most widely used, implement them and standardise them so as to obtain data that is as homogeneous as possible and allows for consistent comparisons. Second, new research methodologies must be developed to overcome a too rigidly binary approach and study diversity in audiovisual professions. More precisely, further qualitative research is needed to enlighten the relations between sex/gender identity/sexual orientation and how these aspects impact the working conditions in the audiovisual sector.

BIBLIOGRAPHY

- Amaturo, Enrica and Gabriella Punziano (2016). *I mixed methods nella ricerca sociale*. Roma: Carocci.
- APA and Fondazione Symbola (2022). “L’occupazione nell’audiovisivo e la presenza femminile”. In *4° Rapporto sulla produzione audiovisiva nazionale*, 38-41. <https://ricerche.apaonline.it/website/wp-content/uploads/2023/04/4-RAPPORTO-APA.pdf> (last accessed 14-01-23).
- Bielby, Denise D. (2014). “Gender Inequality in Culture Industries.” In *The Routledge Companion to Media and Gender*, edited by Cynthia Carter, Linda Steiner and Lisa McLaughlin, 137-46. Abingdon-New York: Routledge.
- Bondebjerg, Ib, Eva Novrup Redvall, Rasmus Helles, Signe Sophus Lai, Henrik Søndergaard and Cecilie Astrupgaard (2017). *Transnational European Television Drama*. Cham: Palgrave Macmillan.
- Buonanno, Milly (2012). *La fiction italiana. Narrazioni televisive e identità nazionale*. Roma-Bari: Laterza.
- Corbetta, Piergiorgio (2015). *La ricerca sociale: metodologia e tecniche. Vol. II: Le tecniche quantitative*. Bologna: il Mulino.
- Creswell, John W. and Vicky L. Plano Clark (2011). *Designing and Conducting Mixed Methods Research*. Thousand Oaks (CA): Sage.
- D’Amelio, M. Elena and Valentina Re (2021a). “Neither Voiceless nor Unbelievable: Women Detectives & Rape Culture in Contemporary Italian TV.” *MAI: Feminism & Visual Culture* 7. <https://maifeminism.com/neither-voiceless-nor-unbelievable-women-detectives-rape-culture-in-italian-tv/> (last accessed 14-01-23).
- D’Amelio, M. Elena and Valentina Re (2021b). “Un lavoro inadatto a una donna: protagoniste femminili nella serialità crime italiana.” *Arabeschi, Sentieri selvaggi. Cinema e Women’s Studies in Italia* 18. <http://www.arabeschi.it/15-un-lavoro-inadatto-a-una-donna-protagoniste-femminili-nella-serialit--crime-italiana1/> (last accessed 14-01-23).
- DEA (2016). *Gap & Ciak: I divari di genere nel lavoro e nell’industria audiovisiva: lo stato dell’arte*. https://www.irpps.cnr.it/wp-content/uploads/2022/01/rapporto-GapCiak_definitivo.pdf (last accessed 14-01-23).
- DEA (2018) *Gap & Ciak 2. Secondo rapporto DEA – Donne nell’industria dell’audiovisivo*. <https://it.readkong.com/page/gap-ciak-2-9535425> (last accessed 14-01-23).

- Dresner, Lisa M. (2007). *The Female Investigator in Literature, Film and Popular Culture*. Jefferson: McFarland.
- EAO (2018). *Gender Imbalances in the Audiovisual Industries*. <https://rm.coe.int/gender-imbalance-in-the-audiovisual-industries/1680946134> (last accessed 14-01-23).
- EAO (2021). *Diversity and Inclusion in the European Audiovisual Sector, IRIS plus*. <https://rm.coe.int/iris-plus-2021-en1-diversity-and-inclusion-in-the-european-audiovisual-1680a299b9> (last accessed 14-01-23).
- EENCA (2020). *Gender Gaps in the Cultural and Creative Sectors*. <https://eenca.com/eenca/assets/File/EENCA%20publications/Final%20Report%20-%20Gender%20in%20CCS%20EAC%20with%20Additional%20sections%20AV%20and%20Radio.pdf> (last accessed 14-01-23).
- ERGA SG 4 (2019). *Study on Industry-led Good Practices Related to Gender Diversity in the European Audiovisual Sector*. https://erga-online.eu/wp-content/uploads/2020/01/ERGA_2019_SG4_Report.pdf (last accessed 14-01-23).
- Fanchi, Mariagrazia and Matteo Tarantino (2021). “Has Digital Transformation Impacted Gender Imbalance in Italian cinema? A Data Analysis of Creative Clusters 2004-2016”. In *Digital Transformation in the Cultural and Creative Industries*, edited by Marta Massi, Marilena Vecco and Yi Lin, 162-183. London: Routledge.
- Fontaine, Gilles (2022). *Female Audiovisual Professionals in European TV Fiction Production – 2020 Figures*, EAO. <https://rm.coe.int/female-audiovisual-professionals-in-european-tv-fiction-production-2021/1680a626dc> (last accessed 14-01-23).
- Gates, Philippa (2011). *Detecting Women: Gender and the Hollywood Detective Film*. Albany: State University of New York.
- Geena Davis Institute on Gender in Media (2021). *Women Over 50: The Right to Be Seen on Screen*. <https://seejane.org/wp-content/uploads/GDIGM-Next50-WomenOver50-Study.pdf> (last accessed 14-01-23).
- MiC and Università Cattolica del Sacro Cuore (2022). *Gender Balance in Italian Film Crews. Data and Research Policies*. https://cinema.cultura.gov.it/wp-content/uploads/2022/09/GENDER_EQUALITY_SITO_v4.pdf (last accessed 14-01-23).
- Hansen, Kim Toft, Steven Peacock and Sue Turnbull (2018). *European Television Crime Drama and Beyond*. Cham: Palgrave Macmillan.
- Hoffman, Megan (2016). *Gender and Representation in British “Golden Age” Crime Fiction*. London: Palgrave Macmillan.
- Holdaway, Dominic (2017). “La rete sociale del cinema di interesse culturale”. In *Il cinema di Stato. Finanziamento pubblico ed economia simbolica nel cinema italiano contemporaneo*, edited by Marco Cucco and Giacomo Manzoli, 127-69. Bologna: il Mulino.
- Jiménez Pumares, Marta (2021). *Female Audiovisual Professionals in European TV Fiction Production*, EAO. <https://rm.coe.int/female-audiovisual-professionals-july-2021/1680a38bd0> (last accessed 14-01-23).
- Klein, Kathleen G. (1995). *The Woman Detective: Gender & Genre*. Urbana & Chicago: University of Illinois Press.
- Krijnen, Tonny and Sofie Van Bauwel (eds) (2015). *Gender and Media: Representing, Producing, Consuming*. Abingdon-New York: Routledge.
- Lauzen, Martha M. (2020). *Boxed In 2019-20: Women on Screen and Behind the Scenes*

- in *Television*. https://womenintvfilm.sdsu.edu/wp-content/uploads/2020/09/2019-2020_Boxed_In_Report.pdf (last accessed 14-01-23).
- Liddy, Susan (ed.) (2020). *Women in the International Film Industr: Policy, Practice and Power*. Cham: Palgrave Macmillan.
- Marradi, Alberto (2007). *Metodologia delle scienze sociali*. Bologna: Il Mulino.
- MiC – Osservatorio sulla parità di genere (2022). *La questione di genere tra immaginario e realtà. Primo rapporto annuale dell'Osservatorio per la parità di genere del Ministero della Cultura*. https://media.beniculturali.it/mibac/files/boards/be78e33bc8ca0c99bff70aa174035096/PDF/2022/OssPariGenere/Primo%20rapporto%20annuale_22%20novembre%202022.pdf (last accessed 14-01-23).
- MiC (2020). *Tutti i numeri del Cinema e dell'Audiovisivo*. <https://cinema.cultura.gov.it/wp-content/uploads/2022/03/Tutti-i-Numeri-del-Cinema-e-dellAudiovisivo-%E2%80%93-Anno-2020.pdf> (last accessed 14-01-23).
- Rai(2022). *Monitoraggio sulla rappresentazione della figura femminile, sulla capacità di garantire il pluralismo di temi, soggetti e linguaggi e contribuire alla creazione di coesione sociale nella programmazione Rai trasmessa nell'anno solare 2021*. https://www.rai.it/dl/doc/2022/06/28/1656409866777_Monitoraggio%20Figura%20femminile%20Pluralismo%20sociale%20e%20Coesione%20sociale%20Rai%202021_Analisi%20dei%20contenuti.pdf (last accessed 14-01-23).
- Re, Valentina and Marica Spalletta (2023). “Unsuitable Jobs for Women: Women’s Behind-the-Scenes Employment and Female On-Screen Representation in Italian TV Crime Drama.” *Comunicazioni sociali* 1: https://doi.org/10.26350/001200_000177.
- Reddy, Maureen T. (2003). “Women Detectives.” In *The Cambridge Companion to Crime Fiction*, edited by Martin Priestman, 191–207. New York: Cambridge University Press.
- Tepper, Daphne (2020). *Achieving Gender Equality and Promoting Diversity in the European Audiovisual Sector: Good Practice Handbook*. https://www.equalitydiversityinavsec.or.eu/pdf/handbook_en_web.pdf (last accessed 14-01-23).
- Turnbull, Sue (2014). *The TV Crime Drama*. Edinburgh: Edinburgh University Press.
- Voices of Culture (2020). *Gender Equality: Gender Balance in the Cultural and Creative Sector*. <https://voicesofculture.eu/wp-content/uploads/2020/02/VoC-Brainstorming-Report-Gender-Balance-in-the-Cultural-and-Creative-Sectors.pdf> (last accessed 14-01-23).
- WIFTMI (Women In Film, Television & Media Italia) (2019). *Women Directors and Film Production: Addressing the Gender Gap and the Role of Festivals*. <https://wiftmia.lia.it/wp-content/uploads/2021/12/WIFTMI-VENEZIA-2019-Women-Directors-Research.pdf> (last accessed 14-01-23).

WOMEN BEHIND THE SCENES: A DATA-DRIVEN APPROACH TO ANALYSE FEMALE LABOUR IN ITALIAN TELEVISION



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.d09926a5>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

5. Constructing an Open, Participatory Database on Gender (In)Equality in the Italian Film Industry

Methodological Challenges

Rosa Barotsi, Mariagrazia Fanchi and Matteo Tarantino¹

ABSTRACT

Through CENTRIC (CrEative NeTwoRks Information Cruncher), an online search tool developed by the Digital Humanities Lab at the Catholic University of Milan, data was gathered by the authors on ten professional roles in Italian feature film productions from the institution of Legge Corona in 1965 and until 2020, with plans to update its findings year by year. The aim was to understand the development of gender equality dynamics in the film industry across this nearly 60-year timespan. Developed in the context of the Horizon2020 project CineAF, the CentricWEBI/CineAF database wishes to serve as a valuable resource for those who are intent on promoting gender equality in policy, film programming and educational curricula. As part of this goal, the ambition is for the database to be open and participatory. A Pilot project was conducted in Spring 2022 in order to test the usability of the platform, and its capacity to generate knowledge about women professionals and reflections on gender inequality. The article focuses on some of the challenges and ethical issues that were made evident thanks to this preliminary exercise, and suggests some possible solutions. These challenges include the ethical implications related to the collection of gender data; the meaningful application of FAIR principles in the humanities and social sciences; and the epistemological pitfalls of film industry data.

KEYWORDS

Gender equity; big data; knowledge ecology; audiovisual industries; film policy.

¹ Rosa Barotsi (University of Modena and Reggio Emilia) – rosa.barotsi@unimore.it; Mariagrazia Fanchi (Catholic University of Milan) – mariagrazia.fanchi@unicatt.it; Matteo Tarantino (Catholic University of Milan) – matteo.tarantino@unicatt.it. All authors contributed to the writing of this article. Rosa Barotsi wrote pages 99-102, Mariagrazia Fanchi wrote pages 88-93, and Matteo Tarantino wrote pages 94-99.

Introduction

What gender equality initiatives are possible and probable depends on what gender knowledge is produced and circulated, and which aspects of gender inequality that knowledge makes visible.

– Eikhof et al. 2019: 841

As Eikhof, Newsinger, Luchinskaya, and Aidley suggest in their analysis of gender inequality in the British screen industries, the availability of data on the gender gap is a necessary condition for planning actions that are able to generate real impact, but it is not a sufficient one. Indeed, as the authors highlight, how data are collected and communicated influences the possibility of launching effective policies supporting gender equality. In the UK, for example, at the time this essay was being written, the social discourses on inequalities tended to focus on the scarcity (or absence) of women in top roles,² generating an “ecology of knowledge” strictly focused on the numbers and percentages estimating the presence of women in different professional roles – and especially in the decision-making ones, but much more imprecise when it came to describing their working conditions, with the consequence of pushing political intervention more towards increasing the number of women employed in the screen industries than towards protecting them: increasing their salary, making their jobs more stable, improving the balance between professional and personal life, and creating the best conditions for overcoming the gender gap, starting from education and training paths.³

² There are, however, some important exceptions to this trend, such as the ‘Raising Films Project’. <https://www.raisingfilms.com/about/> (last accessed 13-04-23).

³ Iñaki Permanyer had carried out a similar analysis, some years earlier, on the criteria used by the European Institute for Gender Equality to measure the gender gap, show-

This paper seeks to problematise some aspects of the relationship between the “ecology of knowledge”, data, and gender equality policies, starting from the CENTRIC project.

CENTRIC (Creative Networks Information Cruncher) is a software developed by the Digital Humanities Lab of Università Cattolica in 2018.⁴ CENTRIC automatically collects and analyses data relating to Italian films – such as year of production, format, genre, production companies, and year of distribution. It also searches and archives data on film crews – at this time concerning ten primary professions: director, screenwriter, editor, producer, music director, director of photography, set design, costume and make-up supervisor, and special effects director –, used to calculate a set of concise indexes devoted to evaluating the gender balance (for instance the Gender Balance Score – Fanchi 2021; Fanchi, Tarantino and Barotsi forthcoming) and describing the level of integration of male and female professionals in the Italian film industry (for instance the gd-score – Fanchi and Tarantino 2019, 2021).

Following Eikhof et al. (2019), in order to understand the meaning and the value of the data generated by CENTRIC, and the representations of the gender gap in the Italian screen industry that it produces, it is essential to understand the specific “knowledge ecology framework” (Eikhof et al. 2019: 841) within which CENTRIC was developed. In other words, as for any other method and scientific approach, CENTRIC, and the kind of knowledge it produces, reflects a set of conditions and coordinates, historically contextualised, that have contributed to defining its priorities, focusing on specific aspects of the complex issue of gender balance rather than others, and that have influenced its solutions, including methodological and technical ones. In the next section, we will define the main theoretical and cultural coordinates that led to CENTRIC.

ing that the indexes used by the Institute tended to overestimate the gender gap in the countries whose economies were less developed and, conversely, underestimate it in the countries with stronger economies (Permanyer 2015).

⁴ <https://almed.unicatt.it/almed-ldh-lab-digital-humanities-fou-main-4970> (last accessed 26-03-23). CENTRIC was developed in 2018 by Matteo Tarantino and it was first used, by two of the authors, to collect and analyze the crews of 3542 Italian films (both shorts and feature films) that received, from 2004 to 2016, public funding as ‘cultural interest films’ (‘film di interesse culturale’) under the so-called Urbani Law (DL 28/2004) (Fanchi 2020, Fanchi and Tarantino 2021).

The CENTRIC Framework

The development of CENTRIC has been influenced by a complex set of factors: cultural, social, legislative and scientific.

In the second decade of the new millennium, the awareness of the profound gender inequalities in the cultural and creative industries, even in the most economically advanced countries (Conor et al. 2015), grew for both policymakers and scholars. Along with the commitment of the European Union and the Council of Europe on gender equality issues, this led to more systemic approaches to the analysis of the presence and roles of women in the screen industries. The ambition was not only to provide a snapshot of gender inequality, but also to promote pluralism and inclusiveness where culture is produced and to reflect on methods and tools that can overcome the absence of data on women's work. The tendency of economic systems to relegate women to the 'grey area' of the labour market, with temporary, informal or 'labile' contracts (Morgan and Nelligan 2015), in fact, makes the presence and contribution of female workers challenging to recognise. This is also because there is a lack of categories suitable to describe and detect the specific forms that women's work assumes. Barbara Montesi wrote:

Statistical survey, which has been and is, a source of primary importance for verifying the dimensions of women's access to different professions and their evolution over time has contributed to the codification of the bias for the 'awkwardness' often shown by statistical surveys in framing the multiple activities of women (Montesi 2002: 148).

In that climate, the new data-driven approaches seemed to represent a veritable turning point; they offered, in fact, the possibility to disclose otherwise invisible correlations, shedding light on the patriarchal mechanisms that still operated in advanced economies and to identify new and more perspicuous categories capable of representing women's work. Mark Lutter's well-known study of the casts of 350,000 U.S. film productions between 1929 and 2010, for example, was welcomed by the scientific community as a truly innovative approach. Using data-driven methods and, especially, Social Network Analysis, Lutter demonstrated that the different career possibilities for the actresses and actors employed in Hollywood cinema depended on the kind and 'quality' of the professional relationships they were able to activate. For instance, Hollywood actresses had less linear job trajectories than their male peers and significantly higher dropout rates when they were

working in ‘closed’ creative teams composed of professionals who tended to collaborate exclusively with each other. Conversely, the risk of failure and career interruption was lowered for actresses involved in projects with ‘open’ creative teams.

Lutter’s essay showed the relevance of ‘relationship capital’ as a resource for female professionals against segregation and gender inequality, highlighting the opportunities of an alternative methodology to the simple counting of men and women employed or represented in film or television (Verhoeven et al. 2020). It also demonstrated the fruitfulness of an approach that did not focus exclusively on what was happening onscreen but on the complex offscreen dynamics.

The focus on the offscreen dimension, at that time, was also supported by a general shift in the scientific paradigm and by the turn from the idea of cultural production as an individual initiative to a collaborative model, typical of digital cultures (Hartley et al. 2015); in this perspective, the absence of women or talents belonging to socially or historically minoritised groups in production initiatives was presented not only as a problem of social justice but as a weakness in the functioning of the media system, with detrimental effects on the cultural and creative industries (Graham and Gandini 2017). The persistence of discriminatory mechanisms, in fact, undermined the principle of ‘everyone’ – which, according to Hartley, Wen and Li (2015), is one of the new millennium’s creative industries’ three pillars, alongside the principles of ‘everywhere’ and ‘everything’, on which their ability to be innovative and generate value depends.

The CENTRIC project took shape at this particular conjuncture, also prompted by the approval at the end of 2016 of the long-awaited new Law on Cinema and Audiovisuals (Law 220, also known as the Franceschini Law), that replaced the previous law, approved in 1965 (Law 1234, known as the Corona Law), and by the absence, in the text of the Law, of any reference to gender equity policies (Fanchi 2020). Even if, in the following months, a first set of incentives for gender equality were introduced, through some implementing decrees (07/31/2017 and 11/09/2017), at the time of its approval, the Law appeared inexplicably indifferent to the issue of the gender gap.

The CENTRIC project took shape within this framework, setting itself three main objectives:

1. establish a sustainable continuous survey system capable of recording women’s presence in the Italian cinema and audiovisual sector and its changes;

2. extend the focus to a range of professions beyond director or screenwriter, to trace the contribution of female professionals to technical roles as well;
3. focus on relationships, measuring the quality and duration of collaborations and more generally the working conditions of women and men in the Italian cinema and audiovisual sector.

However, during the development of the software devoted to data collecting and management, CENTRIC was faced with some new challenges. These challenges were connected, on the one hand, to the quality and completeness of the available information on the workers employed in Italian cinema; and, on the other hand, to the crucial issue of the automatic gender attribution to professionals, and the ethical and political matters it raises. Moreover, the scientific and disciplinary debate, after an initial moment of uncritical enthusiasm for the use of big data, began to detect some critical, even profound, issues linked to the use of data-driven approaches in studies aimed at assessing and overcoming gender inequalities (D'Ignazio and Klein 2020). In particular, with specific reference to the application of such approaches to the field of cinema and audiovisuals, two main issues emerged: the risk of a regression to a biologically deterministic and binary approach, even within frameworks that do not embrace the gender binary perspective (Eikhof et al. 2019, Guyan 2021); and the inability of data-driven approaches to reflect intersectional dynamics (Cullen and Murphy 2018, Jansson and Wallenberg 2020).

Two additional facts helped direct the project and define its particular 'ecology'.

The first one was the beginning of a collaboration with the Italian Ministry of Culture for the assessment of public aid to cinema and audiovisual media. In 2019, Università Cattolica, together with the consulting firm PTSCLAS, was awarded the tender for the analysis and evaluation of the impacts of aid provided by the Italian State to screen industries within the framework of Law 220/2016.⁵ The evaluation model included, in addition to the analysis of industrial and economic effects of public funds

⁵ "Valutazione di impatto", *Direzione Generale Cinema e Audiovisivo* <https://cinema.cultura.gov.it/comunicazione/eventi-e-pubblicazioni/valutazione-di-impatto/> (last accessed 26-03-23).

to the sector, the analysis of social and cultural impacts too, including the ability of the State to encourage, through its various measures, the overcoming of the gender gap in screen industries. This project led to the development of specific sections dedicated to the topic within the Annual Report of ‘Valutazione di Impatto Legge Cinema’, and to the establishment of the ‘focus’: ‘Gender Balance in Italian Film Crews. Data and Research Policies’, an annual Report, first presented in 2021 at the Venice Film Festival, aimed at measuring gender balance in Italian production initiatives, film and audiovisual works, that received clearance for distribution in Italy from 2017, when the new Law came into force.⁶

The second variable, which intervened significantly in CENTRIC’s ‘knowledge ecology’, was the launching, in 2020, of the project ‘A Cinematic Archive for the Future: Women’s Films in Italy, 1965-2015’ (CineAF). CineAF was funded by Horizon 2020 under the Marie Skłodowska-Curie Actions grant.⁷ It was devoted to making visible the contribution of women professionals to the history of Italian cinema, creating a broader awareness that could inspire interventions, in terms of policy and beyond, for the promotion of gender equality in the audiovisual field.

The meeting between CENTRIC and CineAF allowed for an improved approach to gender imbalance in Italian cinema, developing a methodology nestled within the emergent data-based research framework and combining quantitative and qualitative methodologies with feminist research principles and a humanities perspective.

⁶ “Focus parità di genere”, *Direzione Generale Cinema e Audiovisivo* <https://cinema.cultura.gov.it/per-gli-utenti/approfondimenti-tematici/focus-parita-di-genere/> (last accessed 26-03-23).

⁷ The project took place between 2020-2022 and was conducted by Rosa Barotsi under the supervision of Mariagrazia Fanchi, with the collaboration of Matteo Tarantino. One of CineAF’s objectives was to investigate the presence and work of women directors in the Italian film industry between 1965 and 2015, although this was eventually extended to the present. Specifically, at least at this stage, the research focused on women-led feature films that obtained clearance (nulla osta) for national theatrical distribution. <http://cineaf.eu/> (last accessed 04-04-23). This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant agreement No. 891966.

CENTRICwebi

Starting from these premises and taking advantage of the favourable conditions described above, the CENTRICwebi website was launched in 2021.⁸

CENTRICwebi makes available data extracted from CENTRIC and archived in CineAF, allowing it to be queried through a system of filters and visualizations. CENTRICwebi includes two main domains of data:

- Data regarding the audiovisual works, including title, unique identification, production companies, year of production, length and format, nationality, genre, as well as a variety of scores (mostly on gender balance) calculated by the system; for a part of the corpus, the dataset also includes box office and production costs.
- Data on the production crews, focusing on ten professional roles: director, screenwriter, producer, editor, director of photography, heads of department in music, costumes, makeup, set design, and Special Effects, and including such information as the number of professionals in each role, name/surname, and gender; the latter is automatically assigned by the software based on the professional's first name. We will return to this point, which is decisive for the meaning and value of the CENTRIC/CineAF 'ecology of knowledge'.

As of October 13, 2022, the database contains 132,882 entries for professionals, corresponding to 47,554 individuals.

ATTRIBUTED GENDER	% OF TOTAL
Company	21%
Female	21%
Male	56%
Other/Unknown	2%
TOTAL	100%

TABLE I

Distribution of professionals in the database by attributed gender.

Source: CENTRIC/CineAF. Università Cattolica del Sacro Cuore, Milan.

⁸ Like the CENTRIC software, the CENTRICwebi platform was created by Matteo Tarantino. The intellectual property of the platform was acquired by Università Cattolica del Sacro Cuore. The platform is currently in beta version, and it can be accessed at the following address: <http://80.211.232.48:8000/home/> (last accessed 04-04-23).

As of 2023, therefore, the CENTRIC/CineAF project evolved into a tool with the following characteristics. First, it continuously monitors the crews of Italian films and audiovisual works, which received clearance for distribution (Law 161/1962) and, since November 2, 2021, the ‘classification of works for the protection of minors’ (Law 220/2016), from 1964 to the present.⁹ It extracts and cross-references the information from multiple movie databases. The data are automatically updated twice a month. At this point it automatically attributes gender according to binary classification. As gender information is not included as a field in online databases, genders are inferred through probabilistic estimates. Such estimates consider the given names and, when available, short biographies. We chose not to process images, even when present, for ethical reasons. Thus, the resulting “attributed gender” should be considered not as the gender of the *individual* but as that of the *name* (and pronouns used in the available biographies) that the individual employed in the credits of the work. We must keep in mind the scientific, ethical, and political limitations of this approach. At the same time, this raises the methodological question of what social indicators are available to perform any sizeable data-driven analysis of gender dynamics. We will return to this point in the last section. Finally, the system aggregates data into synthetic indexes, aiming to offer a more powerful lens for observing and understanding the features that discrimination based on gender takes in the Italian screen industries, reasoning not only in terms of numerical or percentage presences. Gender Balance Score (GBS) is the most important of these indexes. The GBS is calculated for each profession in each work; it measures the ratio of male professionals to females, ranging from -1 when 100% male to +1 when 100% female.¹⁰ In the current design of the system, only such aggregated indexes are visible to the general public.

Data collected in CENTRIC are managed in compliance with the FAIR guiding principles, particularly concerning issues regarding findability,

⁹ Our core official sources proceed from the Italian Ministry of Culture’s distribution/censorship clearance records (<https://cinema.cultura.gov.it/database-nulla-osta-film/>, last accessed 04-04-23) and datasets regarding films applying for public funding (<https://cinema.cultura.gov.it/comunicazione/eventi-e-pubblicazioni/valutazione-di-impatto/>; last accessed 04-04-23).

¹⁰ GBS operates according to the following formula: $(\text{women} - \text{men}) / (\text{women} + \text{men} + \text{unidentified})$. There is no qualitative dimension related to positive or negative values.

re-usability, and accessibility; all data can be downloaded and reused.¹¹ Furthermore, CENTRIC was conceived as a participatory archive. This choice has a twofold purpose. The first is to make the project independent of the researchers who launched it and thus enable it to continue over time, interweaving other ‘knowledge ecosystems’, and, in this way, improving the quality of data and their ability to inspire policies and actions capable of effectively supporting gender equality. The second reason concerns the nature of the data and the imprecision and lacunae of some information domains. As mentioned at the beginning of this paper, information about film workers is public and accessible but less complete and correct than one might imagine. The participatory approach allows for correcting these errors, integrating missing information, and acting on the crucial issue of automatic gender attribution.

Among the many issues we have addressed during the development of the project and dataset, we would like to focus on two specific points: the exhaustiveness of the data and the crucial issue of automatic gender attribution to professionals.

A Trace of the Instability of Online Film Information: The ‘Pilot Experiment Initiative’

Despite the abundant information about films (at least from the institution-alisation phase of cinema onward), cinema configures a field of “dirty data”: different countries produce fragmented institutional datasets of various types of films, rarely complete and mostly related to public funding; access to databases is slow and heavily regulated; there exists a plethora of private initiatives (including Amazon’s IMDB) that try to fill the void, offering partial, non-comparable and sometimes contradictory information. Especially, the further one moves away from Hollywood, the more inaccurate cinema becomes as a domain of data, even with reference to basic master data, such as the title: titles can be multiple, of course, in different languages, have different spelling, and so on.

¹¹ At the time this text was being written, a set of ethics guidelines and rules were being drafted to protect sensitive data, as required by law, and to ensure that the epistemological limits of the dataset (especially regarding automatic gender attribution) are clear to any users outside the research team.

To assess the completeness of the data collected automatically by CENTRIC and thus provide a measure of the (possible) inaccuracy of the available public cinema data, between February 22, 2022, and May 31, 2022, the *Pilot Experiment initiative* was launched.

During this Pilot Experiment, 267 editors, operating under the supervision of two of the authors, worked on 15 selected years (1964-1967 and 2008-2020, for a total of 3944 works or 34.5% of the total works and corresponding to 39% of the crew entries) manually checking the information collected.

Editors were assigned a package of works and manually double-checked all information entered with respect to those works and the corresponding crew by examining available information (such as biographies, interviews etc.). Each edit had to be manually approved by a supervisor.

This resulted in a total of 21,557 edits performed on the database. Crew edits (including modification, insertion, and deletion) amounted to 82.18%; these modifications accounted for 28.01% of the total. Thus around 36% of the crew information in the sample required corrections.

Gender was the core concern of the project underpinning CENTRIC. We expected a significant instability of gender information. This was not confirmed by the data. The overall rate of professionals whose attributed gender information was modified is the same for both attributed gender categories, at around 3.6%. Thus about 96.4% of the subsample had the appropriate gender attribution based on the available information and according to our method. However, we must always keep in mind that gender values could only be expressed through the Male/Female dichotomy and through an imperfect proxy, as the available information currently does not allow for efficient estimation through other options.

ATTRIBUTED GENDER	GENDER EDITS	TOTAL	% GENDER EDITS
Female	195	5,219	3.7%
Male	532	14,805	3.6%

TABLE 2

Gender distribution of professionals in the database whose attributed gender information was edited.

Source: CENTRIC/CineAF. Università Cattolica del Sacro Cuore, Milan.

Edits in themselves must be coupled with analyses of rejection rates, that is, the number of times a typology of edit was refused by the supervisor. In our case, gender showed a low degree of ambiguity with a rejection rate around

1%, whereas edits on profession and film title were rejected at much higher rates, peaking at 5.34% for film titles. However, in our case, the value of this data is severely skewed by an initial edit policy that limited gender edits to supervisors, due to the sensitivity of the information. Thus the overwhelming majority of gender edits were done by supervisors themselves, who tended to be well aligned and not to reject each other's edits.

At any rate, the instability of film titles shown by rejection rates should not be discarded as mere happenstance. Film titles are important identifiers, often used (also by CENTRIC) as a core index variable for cross-referencing film information proceeding from different data sources. In other words, we often recognize two works as the same because they share the same title.

As each database uses its own unique identifier for professionals and works, it is impossible to reference them across sources outright. Yet using titles for index variables runs into further complications: not only may there be films with the same titles released in the same period (this is particularly true for small independent films), but titles of the same works might be reported by different sources with differences small (for instance using apostrophes instead of accented letters) and large (for instance, changing the title completely from preproduction phase to the final release).

In general, index variables are the ones showing the higher instability in our dataset. Most of the crew modification edits regarded alignment of IDs of professionals or works, amounting to 78% of total modifications (see Table 3). This means the system failed to probabilistically attribute entries to the same professional (most notably because the professional's name was reported differently across sources, or because filmographies showed gaps – for instance for early-stage professionals) or professionals to the same work.⁴

FIELD MODIFIED	% OF TOTAL
Prof ID	64%
Film ID	14%
Gender	9%
Film Title	5%
Name	4%
Job/profession roles	4%

TABLE 3

Percentage of edits by field modified.

Source: CENTRIC/CineAF. Università Cattolica del Sacro Cuore, Milan.

While this, of course, points to the need to strengthen CENTRIC's probabilistic match algorithms, it also points to more significant underlying problems in index variables – and with them, of the stability of online data about Italian cinema. The work of alignment recorded in our databases is driven by the lack of unique, stable identifiers across the multiple available databases regarding both works and professionals. This significantly raises the operational cost of complex, big-data-driven analyses of the cinema industry.

Ethical and Methodological Questions Relating to Gender Attribution

As mentioned earlier, collecting data on gender identities that fall beyond the binary is a complex question for big-data research. If someone might interject that the number of people who fall outside the dimorphic genders is too small to matter, the answer would be obvious: it is precisely this exclusionary feedback loop that leads to such individuals being erased from data in the first place (Lindqvist et al. 2020: 332). In other words, we must be aware that data proceeding from a binary gender dichotomy may be trading in accuracy (and even discrimination) for efficiency. Additionally, until we find and implement an appropriate way to gather information on gender-variant people, we will not know the scale of that inaccuracy. For example, in the data we collected during the summer of 2021 from the Ministry of Culture on distributed feature films between 2017 and 2020, we made use of the gender binary to check for any progress made in the employment situation of women since the introduction of gender parity incentives through the new law. This meant that we may have failed to account for filmmakers such as Giuseppe Sciarra who, as we were in the process of data collection, came out publicly as non-binary.¹² On the other hand, if anonymization processes are less than perfect, the collection of data on gender-variant identities runs the serious risk of revealing information about people's gender without their consent. To speak of gender variant

¹² “Giuseppe Sciarra – ‘Non mi sento né uomo né donna’” 2014. <https://cinemaitaliano.info> (last accessed 26-03-23). In this particular case, our data collection was limited to feature films, so Giuseppe Sciarra would have not been part of our dataset since their recent output consists in short films. Sciarra's case is indicative, however, of the types of exclusions we are likely to operate if we continue to use the gender binary in data research.

film professionals in this context complicates the principle of “making the invisible visible”, if the terms under which that visibility is performed are not constantly negotiated with the people who are being made visible. In any case, it is important to remember that quantitative work on groups that have been systematically excluded involves “carrying a huge imbalance of power”, including the power to misgender or out people against their will (Matias 2014).

This is one of the challenges we came up against when deciding on our own strategy for data collection. Because we are partly dealing with historical archival data about thousands of people, a significant percentage of whom are no longer alive or traceable, we made the decision to assign gender based on first names and following the gender binary. This is far from ideal. Although we tried to guard against some of the problematic aspects of this method (as we will describe momentarily), it would be disingenuous to claim it is impossible to check with each and every person of a larger dataset to make sure their gender is represented correctly: a more honest explanation is that it is *too slow*, and therefore more labour – and resource – intensive. Especially in an academic environment that demands fast turnover of research results – both as a success fetish and as a material condition resulting from precarious, short-term research contracts – it is difficult to advocate for research methods that appear less than efficient. Such dilemmas bump up against a basic premise of feminist ethics – the fact that it takes more time.¹³

To return to our project, in order to counteract some of the problematic aspects of our chosen method, we make it clear that gender is assigned by the system and in reference to the name, not the person. To the extent that we have been able to manually check a sample of the dataset, the automatically-assigned gender was cross-referenced with official accounts of the film professionals, in magazine or newspaper articles, and, wherever possible, interviews or personal websites in which we hear their own voice, in order to determine how they self-identify publicly (for instance, through the use of pronouns or adjectives, which are gendered in the Italian language).

Although still imperfect, this method allows us to establish that these professionals were socially perceived, or presented as, women, and were

¹³ This is discussed, for instance, by the US-based filmmakers collectives Ethnocine and Rhiza in Episode 1 of their podcast, “Bad Feminists Making Films.” <http://badfeministmakingfilms.com/podcast> (last accessed 26-03-23).

therefore subject to the same biases and systemic exclusions as other women of their time. By assigning a gender based on the traditionally accepted gendered meanings of first names, in other words, we understand “woman” as an active category of social signification. Gender therefore functions not as a reflection of the gender identity of the individual, but as a term that condenses the norms and regulations that a patriarchal society applied to that individual.¹⁴

What this means, as already noted, is that we don’t have data on gender categories outside the binary. To make sure that this is understood as a lack in our research, we intend to note this absence in our text and figures. In the next steps of the project, we will integrate our approach to data collection with case studies that provide deep and textured stories of individual creators, to restore some of the context that accompanies our data.

Finally, and most importantly, the participatory nature of CENTRICwebi means that the professionals themselves will be able to make modifications and integrations to the collected data, including data on their gender, across their entire filmography or for specific productions, thanks to the dataset’s molecular approach (Fanchi, Barotsi and Tarantino forthcoming)

Many questions remain open: in order to render the CENTRIC/CineAF database open and participatory, one of the fundamental next steps is to establish a set of ethics guidelines relating to access and verification. This will undoubtedly involve complex decision-making. But, as with previous stages of the project, navigating those challenges as an interdisciplinary research team becomes part of the knowledge production process, and sharing them with other researchers in contexts such as this one is, as I hope you will agree, crucial.

¹⁴ Natalie Wreyford and Shelley Cobb follow a comparable logic, in that they suggest they attempted ‘to “read” a person’s gender identity the way a potential employer would’ (2017). Nancy López (2014) has developed the terms “street race,” “street gender,” and “street race-gender” as tools for use in the US census, with the capacity to separate between a person’s self-identity and the way other people, and therefore institutions and decision-makers, perceive them based on societal bias and contextual norms. A question she suggests might concisely account for the distinction between gender and ascribed gender as a census question might look like this: “If you were walking down the street, how would other Americans who do not know you personally identify your gender?”

Conclusions

CENTRIC defines a field of knowledge oriented towards providing solutions to the issue of the gender gap in the specific historical moment and framework in which the researchers found themselves operating. In this sense, as any other research project, it is physiologically limited and susceptible to additions, modifications, extensions, and changes. Additionally to those planned improvements that have already been mentioned, concerning, for instance, the question of automatic gender attribution or the sharing of ethical and political assumptions of knowledge, there are others with which researchers are now faced. For instance, the focus on the off-screen dimension, that characterised the early stages of CENTRIC's work, prompted by the need to compensate for the absence of data on gender balance in crews, requires an extension to the dimensions of onscreen representation and access to the cultural experience of spectatorship for all people. This expansion from the offscreen to the onscreen and in-front-of the screen, as well as the challenge of applying intersectional parameters to the discrimination mechanisms in screen industries, requires a never-ending redefinition of the perimeters and coordinates of CENTRIC's 'knowledge ecology', also by creating synergies with other projects that cover complementary areas.¹⁵

We are convinced that, in this way, it is possible to bring 'knowledge ecologies' to life, constantly evolving to reflect the needs and challenges of the present, and capable of offering an effective and powerful tool for the promotion of inclusiveness and cultural pluralism.

¹⁵ For example, the University of Bologna has developed a series of tools to assess gender balance in representations, with a focus on the imagery of TV series (Rocchi et al. 2023, Rocchi 2023).

BIBLIOGRAPHY

- Conor, Bridget, Rosalind Gill and Stephanie Taylor (2015). "Gender and Creative Labour." *The Sociological Review* 63(1 suppl): 1-22. <https://doi.org/10.1111/1467-954X.12237>.
- Cullen, Pauline and Mary P. Murphy (2018). "Leading the Debate for the Business Case for Gender Equality, Perilous for Whom?" *Gender, Work & Organization* 25(2): 110-26. <https://doi.org/10.1111/gwao.12199>.
- D'Ignazio, Catherine and Lauren F. Klein (2020). *Data Feminism*. Cambridge: The MIT Press.
- Eikhof, Doris Ruth, Jack Newsinger, Daria Luchinskaya and Daniela Aidley (2019). "And ... Action? Gender, Knowledge and Inequalities in the UK Screen Industries." *Gender, Work & Organization* 26(6): 840-59.
- Fanchi, Mariagrazia (2020). "Equality Illusions. Finanziamento pubblico del cinema e politiche di 'genere'." In *Il sistema dell'impegno nel cinema italiano contemporaneo*, edited by Dom Holdaway and Dalila Missero, 187-203. Milano: Mimesis.
- Fanchi, Mariagrazia (2021). "Equilibri ed equilibrismi. Bilanci di genere nella produzione cinematografica e audiovisiva italiana. 2017-2020." *8 e 1/2. Numeri, visioni e prospettive del cinema italiano* 60(12): 68-73.
- Fanchi, Mariagrazia and Matteo Tarantino (2021). "Has Digital Transformation Impacted Gender Imbalance in Italian Cinema? A Data Analysis of Creative Clusters 2004-2016." In *Digital Transformation in the Cultural and Creative Industries*, edited by Marta Massi, Marilena Vecco and Yi Lin, 162-84. London and New York: Routledge.
- Fanchi, Mariagrazia and Matteo Tarantino (2019). "Donne e reti creative nel cinema italiano (2004-2016)." *Economia della Cultura* 4: 523-30.
- Fanchi, Mariagrazia, Matteo Tarantino and Rosa Barotsi (2023). "Heaven Can Wait? Gender (Im)Balance in Contemporary Italian Film Crews." *Comunicazioni sociali* 1: 98-113. https://doi.org/10.26350/001200_000178
- Graham, John and Alessandro Gandini (eds) (2017). *Collaborative Production in the Creative Industries*. London: University of Westminster Press.
- Guyan, Kevin (2022). "Constructing a Queer Population? Asking about Sexual Orientation in Scotland's 2022 Census." *Journal of Gender Studies* 31(6): 782-792. <https://doi.org/10.1080/09589236.2020.1866513>.
- Hartley, John, Wen Wen and Henry Siling Li (2015). *Creative Economy and Culture. Challenges, Changes and Futures for the Creative Industries*. London: Sage.
- Jansson, Marie and Louise Wallenberg (2020). "Experiencing Male Dominance in Swedish Film Production." In *Women in the International Film Industry*, edited by Susan Liddy, 163-78. Cham: Palgrave Macmillan.

- Lindqvist, Anna, Marie Gustafsson Sendén and Emma A. Renström (2021). “What Is Gender, Anyway: A Review of the Options for Operationalising Gender.” *Psychology & Sexuality* 12(4): 332-44. <https://doi.org/10.1080/19419899.2020.1729844>.
- López, Nancy (2014). “What’s Your “Street Race-Gender”? Why We Need Separate Questions on Hispanic Origin and Race for the 2020 Census.” *Robert Wood Johnson Foundation*. https://www.rwjf.org/en/insights/blog/2014/11/what_s_your_street.html (last accessed 26-03-23).
- Lutter, Marc (2015). “Do Women Suffer from Network Closure? The Moderating Effect of Social Capital on Gender Inequality in a Project-Based Labor Market, 1929 to 2010.” *American Sociological Review* 80(2): 329-58. <https://doi.org/10.1177/0003122414568788>.
- Montesi, Barbara (2002). “Donne, professioni, cittadinanza.” *Storia e problemi contemporanei* 31(9): 147-66.
- Morgan, George and Pariece Nelligan (2015). “Labile Labour – Gender, Flexibility and Creative Work.” *The Sociological Review* 63(1 suppl): 66-83. <https://doi.org/10.1111/1467-954X.12241>.
- Nathan, Matias J. (2014). “How to Ethically and Responsibly Identify Gender in Large Datasets – MediaShift.” *MediaShift*. <http://mediashift.org/2014/11/how-to-ethically-and-responsibly-identify-gender-in-large-datasets/> (last accessed 26-03-23).
- Permanyer, Iñaki (2015). “Why Call It ‘Equality’ when It Should Be ‘Achievement’? A Proposal to Un-Correct the ‘Corrected Gender Gaps’ in the EU Gender Equality Index.” *Journal of European Social Policy* 25(4): 414-30. <https://doi.org/10.1177/0958928715594545>.
- Rocchi, Marta, Lorenzo Cattani and Guglielmo Pescatore (2023). “Gender Equality in European Netflix TV Series Production (2014-2019).” *Mediascapes journal* 21(1): 188-204.
- Rocchi, Marta (2023). “Women’s Labour in TV Series Production: A Comparative Analysis of Italian Generalist TV and Pay Platforms (2016-2019).” *Comunicazioni sociali* 1: 69-81. https://doi.org/10.26350/001200_000176.
- Verhoeven, Deb, Katarzyna Musial, Stuart Palmer, Sarah Taylor, Shaukat Abidi, Vejune Zemaityte and Lachlan Simpson (2020). “Controlling for Openness in the Male-Dominated Collaborative Networks of the Global Film Industry.” *PLoS ONE* 15(6): 1-23. <https://doi.org/10.1371/journal.pone.0234460>.
- Wreyford, Natalie and Shelley Cobb (2017). “Data and Responsibility: Toward a Feminist Methodology for Producing Historical Data on Women in the Contemporary UK Film Industry.” *Feminist Media Histories* 3(3): 107-32. <https://doi.org/10.1525/fmh.2017.3.3.107>.

CONSTRUCTING AN OPEN, PARTICIPATORY DATABASE ON GENDER (IN)EQUALITY IN THE ITALIAN FILM INDUSTRY: METHODOLOGICAL CHALLENGES



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.9fa071ea>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

6. Women in Polish TV Series

Qualitative and Quantitative Analysis¹

Andrzej Meler, Beata Królicka, Radosław Sojak

◀ ABSTRACT

Our study concerned the extent to which contemporary Polish TV series tried to change stereotypes about women while maintaining them. The objective was operationalised in terms of the presence and presentation of fictional female characters in series broadcast by the three largest television stations in the Polish television market. Eight series representing four thematic types were monitored. We applied a multi-layered coding to the film material. Quantitatively, we defined the contexts in which female characters were shown more often and less frequently. The series were dominated by contexts related to interpersonal relationships (29%) and relationships of a social nature (26%). Two subsequent contexts were professional work (15%) and crime (10%), but they differed across individual genres and titles. We then conducted a qualitative analysis which consisted of identifying the stereotypes or anti-stereotypes presented in the series.

KEYWORDS

Women's presence; Polish TV series; female characters; gender stereotypes; social change.

¹ The article is an abbreviated and supplemented version of the report of the monitoring commissioned by National Broadcasting Council in 2019. The original report was published as *Stereotypical or Non-typical? Women in Polish TV Series: Their Image, Presence and Context of Appearance* by Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika [Nicolaus Copernicus University Press] in 2020.

Introduction

The experience of all waves of feminism and women's empowerment movements shows that formal (legal) change is only a prelude to informal social change. The latter often occurs many years after the former and is not without additional incentives. This is most likely because every mechanism of real social change is stretched out in a delicate balance between the introduction of new elements into culture and the acceptance of those that already exist (Zybertowicz 1995). If we were to treat this two-stage nature of social change as a guidance for our study, then analysing mere exposure of women in Polish TV series would answer the question about the formal aspect of the change, while analysing how stereotypical female characters are would address the informal aspect. This results in the two-fold structure of the analysis that follows. Firstly, using mainly quantitative methods and measures, we establish the general pattern of the presence of female characters in contemporary Polish TV series. Secondly, moving to qualitative analysis, we will answer the question of to what extent do contemporary Polish TV series try to change stereotypes about women and to what extent do they preserve them.

The Polish Media

The Polish media consists of relatively strong public sector which plays an essential role in balancing the overall political system (Łódzki 2008). In the television sector, the dominant broadcasters are Polish Television [Telewizja Polska] (state-run), Polsat (private, with Polish equity), and TVN (a subsidiary of Discovery in 2018–22). Despite the fact that two broadcasters are controlled by Polish capital, the advantage of domestic capital on the TV market is not as big as it is in Germany or France (Gajlewicz-Korab

and Konarska 2017). Considering the share of foreign capital in the media system, the system has features typical of half-peripheries or peripheries (Radziejewski 2018, Hallin and Mancini 2012).

The national public television has nearly a third of the share of the general viewership (tw 2018). It is also notable that public TV in Poland enjoys the highest geographical and population coverage in the country. There are some portions of territory and population that have only access to public media.

Research Material and Coding Procedure

The research sample comprised of chosen episodes of 8 TV series representing four genres: dramas – soap operas, historical shows, criminal stories and comedies. We had a total of over 49 hours of material to analyse; however, as the opening and closing credits were excluded from the analysis, in further compilations, we consider a sample of 46 hours and 52 minutes. The details of the research sample are presented in Table 1.

GENRE	TITLE OF THE SERIES	BROADCASTER	TIME (HH:MM:SS)	TIME (% OF SAMPLE)
Soup operas	<i>L for Love</i>	TVP	7:11:15	15.33%
	<i>First Love</i>	Polsat	5:03:38	10.79%
Historical dramas	<i>Crown of Kings</i>	TVP	3:47:42	8.09%
	<i>War Girls</i>	TVP	6:47:31	14.48%
Crime dramas	<i>Diagnosis</i>	TVN	6:05:58	13.01%
	<i>Trap</i>	TVN	3:59:20	8.51%
Comedies	<i>Girls from Lviv</i>	TVP	6:52:13	14.65%
	<i>Friends</i>	Polsat	7:06:14	15.15%
TOTAL			46:53:51	100.00%

TABLE 1
Sample metrics.

To attain our research objectives, we had to apply multi-layered (four levels) coding to the film material. While watching individual episodes of the series in a special application, the encoders simultaneously coded specific fragments according to the adopted assumptions with the help of solutions available in the software.

In the visual layer (level one), each team member coded only what was displayed on the screen (omitting the sound) and marked each appearance of a woman, a man, or combinations of the two (two women, two men, a pair, a homogeneous group, a mixed group). Owing to sociological significance, the configuration in which the characters were visible on the screen was divided into three basic types: (1) exposure of a single person; (2) exposure of a pair – dyad (Simmel 2009); and (3) exposure of a group of three or more people. At the same time, we consistently assumed that coding the visibility of a character on the screen requires that the viewer see the face or at least the profile of the character.

In the audio layer (level two), it was necessary to code the duration of the men and women speaking. The encoders listened to the same material and coded only what was heard (omitting the video): a woman's or man's speech, polyphony, or non-speech.

The issue of audio coding seems clear for adults, but it is not entirely clear how to treat children when dividing the genders. This study adopted the general principle that children have gender and are therefore included as women or men in audio and video coding. In special situations, when the child's gender was undetermined (e.g., infant), 'other (including children)' codes appeared and the exposure was not included in the exposure of either women or men, only in the total duration of the recording.

However, the above elements are only an introduction to in-depth image analysis. Visual or sound exposures took place in the series against the background of defined social and cultural contexts. In the third layer, each episode was divided into the contexts in which its action took place.

Besides the above coding, scenes were also cross-sectionally assigned stereotypes (level four). While watching, the encoders marked places in the episodes containing an exemplification of stereotypes.

One could argue that coding for the first two layers (visual and audio) is not a theoretically heavy procedure. The other two (contexts of interaction and stereotypes) are just the opposite and are impossible without elaborate articulation of the theoretical framework.

Theoretical Framework

Contexts of Interactions

One of the key aspects of interpretation of everyday life highlighted in sociology from Erving Goffman (e.g., 1975) to ethnomethodology (Garfinkel 1967) is framing or contextualisation. Words, gestures, and actions receive meanings that are understandable in specific contexts, constituting the basic interpretative resource for both the actors and the audience. One of the basic distinctions for classifying interpretative contexts relates to the division into the private and public spheres. And although this division works well in the analysis of real behaviour, it loses some usefulness in the case of television series. Indeed, genre requirements significantly reduce the exposure to purely private situations. For narrative order, however, let us assume that the division between the private and the public is unclear, and is more of a continuum. This will enable us to organise the contexts of exposure of the characters distinguished for the analysis, from the most public ones, through more personal, to intimate ones. This provides us with the following typology:

1. Professional work
2. Physical and sporting activities in public places
3. Health – mainly in healthcare institutions
4. Social activity – civic activity reaching beyond neighbourly self-help – mainly of a voluntary or quasi-political nature
5. Armed conflict
6. Participation in culture
7. Shopping
8. Criminal activities
9. Sociability
10. Childcare
11. Planning and engaging in intrigues
12. Housekeeping
13. Relationships, feelings, liaisons – cultivating emotional and intimate relationships

Regardless of the above list, the contexts of individual scenes are a variable posing a challenge in analytical terms. Many series show simultaneous or intertwined plots – the same scene can be placed in a professional context and relate to the relationship and emotions of the characters. Coding all the possible contexts included in a scene would lead to assigning it several

contexts from the aforesaid list, which would yield little analytical value. In this situation, we adopted the principle of coding the context which stands out most and determines the characters' behaviour at a given time to the greatest extent.

Stereotypes

Stereotypes are simplified thought constructs generally referring to a community and shared by a group of people. Although the concept itself does not have negative connotations, a large portion of research on stereotypes concerns the issue of prejudice (Boksański 1997), which occurs when an idealised image of one's own community (auto-stereotype) is contrasted with a negative and exaggerated image of a foreign community, e.g., an ethnic or sexual minority (Sztompka 2020: 295).

In the case of gender stereotypes, the simplified image refers to a given gender (women or men) or a subgroup therein. As research has shown, gender stereotypes are characterised by a multilevel structure, within which certain concentrations of stereotyped characteristics can be distinguished. Stereotypes attribute certain characteristics to women and men based on more general categories, such as personality traits, social roles, appearance, and occupation (Mandal 2004: 17-20).

Gender stereotypes in a given culture play an essential role in both individual and social dimensions. They influence the formation of gender identity, and the perception of oneself and one's body. Women, more often than men, display a lack of satisfaction with their own bodies, which in turn results in lower overall self-esteem (Mandal 2004: 35-41, Kochan-Wójcik 2003). The location of the sense of control and the style of attribution are also acquired as per the prevailing gender stereotypes. Compared to men, women are more often characterised by an external sense of control, resulting in a sense of helplessness. Blaming oneself for failures, typical of the feminine attribution style, leads to a lack of self-confidence and a sense of undeserved success. In the process of socialisation, social and professional roles typical of a given gender are assimilated in cultural messages, which influence both the shape of personal life, e.g., friendships, erotic relations (Mandal 2004), and professional career (Domański 1995).

To identify the stereotypical images of women in the series analysed, we used a previously prepared list which includes stereotypes of a different nature and origin. This is because it was composed of various commonplace

stereotypical social roles, personal patterns, cultural myths, or archetypes. All of them were connected by the fact that they portrayed women through the prism of a certain simplified set of traits or by sharpening one dominant trait of character. The list was constructed by the brainstorming procedure of a 10-member team comprising the authors of the study and the team of coders. The majority of the group were women and brainstorming followed individual research on gender stereotypes by each member of the group.

The set of stereotypes used in our research can be classified into those referring to 1) social roles, 2) physical and mental characteristics and sexuality, and 3) professional work.

1. Social roles

- Polish mother/goddess of hearth and home/homemaker – all are closely connected with the idea of home being natural sphere of female activity. Sometimes, it leans on the positive side when stressing the influence of the woman on maintaining the continuity and unity of the family, but is often biased towards negatively implicating an image of a woman constrained within the private sphere, subordinated to a man, restrained, neglected and mentally restricted;
- Matron – usually refers to an experienced mature woman among women or the head of the family;
- Stepmother – non-biological mother. The stereotype of a stepmother is negative, and in line with it, she treats her husband's (partner's) children badly or worse than her own. The stereotype includes both the stepmother's jealousy for the attention of her husband or partner devoted to his children from a previous relationship and the issue of competition for property to which stepsons or stepdaughters are entitled (Jakubiec 2007).

2. Physical and mental characteristics, and sexuality

- greater emotionality – based on a dichotomy of the heart and the mind, a woman is deemed to be more sensitive and tender than a man, who in turn is stereotypically described as more rational (Mandal 2004: 17). In the positive version, a woman's feelings are deeper, she is more sensitive, empathic and more willing to help. In the negative variant, a woman's emotionality may manifest itself as hypersensitivity or border on hysteria.
- *femme fatale* – emphasises both the sexual attractiveness of women and their destructive power and demonic nature. The stereotype re-

fers to a woman who is sexually very attractive and seductive, generally young (though not a teenager). However, it is primarily a woman who leads a man to perdition, through a divorce, loss of property or position, or even death.

- bimbo or blonde – refers to a young woman, physically very attractive, usually with blonde hair, who emphasises her sexual qualities and is treated by men as a sex object. According to this stereotype, physical attractiveness does not go hand in hand with intellect.
- ivy – stereotype emphasising weakness and indecision of women and their dependence on men.
- atechical – refers to the alleged lower proficiency of women in the use of technical devices. This includes the stereotype of a woman as a poor driver (worse than a man) (Chłosta-Zielonka and Matusiak-Kempa 2015) or her inability to carry out small repairs at home.

3. Work

- a successful woman with a ruined personal life – a negative variant of a career woman. Stereotypically, she is a mature (but not old) woman who has paid for her professional success with a miserable private life or has never started a family, is childless and with loneliness as a grave problem (see Czernecka 2011: 129-143).

Stereotypes amalgamate with each other both synchronously and diachronically. From a film art perspective, diachronic connections are an especially important element in building the story, when, for example, in a critical situation, a *femme fatale* turns out to be a lost ivy, and a repressed hausfrau discovers in herself the makings of a strong successful woman. Such measures are important in the context of shaping social change.

Results and Discussion

Active exposure refers to the time spent by men and women speaking, regardless of whether or not the character speaking is shown on the screen. As there is practically no polyphony in the recordings, the active exposure is a disjointed variable, i.e., at the given moment, we can deal with either the active exposure of a woman or a man or a lack of expression (non-speech).

In total, the active exposure of women, i.e., the time women speak, was 17.5 hours, accounting for 38% of the total duration of the action under

analysis. Compared to the active exposure of men, the ratio was 1.19 in favour of women. At the most general level, therefore, we can say that in the series sample studied, women were more often displayed than men, but this was not an overwhelming advantage. Interestingly, nearly one-third are fragments without people's speech, given to music, silence or street noises. These fragments often play an important role in building tension, regulating the pace of action or additionally illustrating the internal states of the characters. Judging by active exposure, we may provisionally conclude that women's share of exposure is at least equal to men's. This picture is, however, incomplete without considering the passive exposure of genders.

Passive exposure means the appearance of women or men on the screen without speaking. It is an indicator methodologically different from active exposure due to the possibility of men and women being shown on the screen simultaneously.

Women are also predominant in passive exposure, with the ratio of their passive exposure to men's being 1.15. In total, the scenes in which women (in any configuration) were shown on the screen (regardless of whether they spoke or not) took 67% of the recording time, while the corresponding share of men was 58%. The remaining 5% of the recordings are taken up by scenes with no humans. Therefore, we can argue that the analysis of passive exposure supports the conclusion suggested above on the advantage of women's exposure in the series analysed – an advantage that is clear and non-overwhelming.

Considering the relationship between the active and passive exposure of women and men, we can conclude that in the case of the two series, we are dealing with a predominance of men concerning both speaking time and screen appearance. *Diagnosis* is a unique series where women dominate over men in the sphere of dialogue, but not on the visual level. Other series show women more often than men to a larger or smaller degree. The ratio of the time of active to passive exposure of women in a given series enables us to assess the extent to which showing women goes hand in hand with giving them the floor. This indicator combines a different series such as *Crown of Kings* and *Friends*, whose active-to-passive exposure ratio is high, i.e., concerning their visibility on the screen, women proportionally speak a lot. In these two, as well as in *First Love* and *Diagnosis*, a woman usually speaks when she is on the screen. At the other extreme, in *War Girls* and *Trap*, a large part of the presentation of women on the screen is silent, which does not mean that they are merely a background but is more a way of building up tension and furthering the plot.

In terms of the active exposure of women, the two comedy series, *Friends* and *Girls from Lviv*, dominate. The other end of this ranking is occupied by the two historical drama series *Crown of Kings* and *War Girls*. However, although the share of the active exposure of women in historical drama series is similar, the relation of this exposure to the active exposure of men is very different. While in *Crown of Kings*, the small active exposure of women is accompanied by a large active exposure of men, probably due to the medieval historical period depicted, in *War Girls*, women speak more than men, though their speaking time is relatively less than in other series. The share of silent scenes is significant in this case.

After the review of active and passive exposures in the series, the question remains as to which series is more masculine and which more feminine, in terms of both exposures. When we combine passive and active exposures, i.e., the audio and video layers, a fairly wide range of configurations is created. We juxtaposed these configurations to provide a spectrum that extends from the configuration in which women are visible on the screen and simultaneously, we hear their utterances to a similar male configuration (Fig. 1). The series with the largest exposure of women are *Friends*, *Trap*, and *Girls*

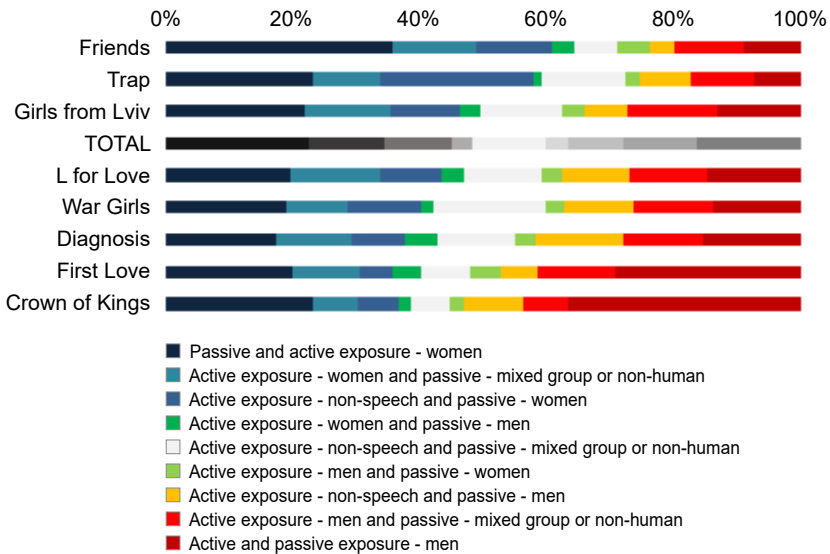


FIGURE 1
 Ranking of the series in terms of total active and passive exposure.

from *Lviv*. The *Friends* series is clearly in a class of its own, when it comes to the time occupied by women both in dialogue and screen visibility. This advantage also has an impact on the average for the total sample tested, clearly elevating it. In the context of *Trap*, it is noteworthy that a large share of women's exposure is passive, i.e., without any speech. This is due to the large number of highly dramatic scenes involving women. In *War Girls*, the sum of exposures involving women is slightly higher than that involving men; but given that the three main characters are female, this does not seem to be a significant advantage. In this series, scenes without dialogue and with the exposure of mixed groups or without human figures have the greatest share of the recording time.

A special case in terms of exposure is the series *Diagnosis*, in which, on the one hand, scenes with women in both passive and active exposure have an advantage over corresponding male scenes, but on the other hand, the greatest share (14%) of the total recording time among all the series comprises scenes in which men appear on video without dialogue. These are, for example, scenes illustrating the characters' inner dilemmas. The large share of such scenes makes *Diagnosis* move to more masculine positions in the exposure ranking.

In turn, the greatest exposure of men is in *Crown of Kings* and *First Love*. These two series clearly expose men much more, which is understandable in the former (medieval historical drama) case and surprising in the latter (classic soap opera).

In light of the above data, we may argue that on a formal level, women are not disadvantaged in terms of exposure in contemporary TV series. On the contrary, they take a majority of on-screen presence (both active and passive) consistently. They have their own voice in the plot. But a closer examination of the various configurations of characters shown brings the first nuances to the above preliminary conclusion.

Women and men are presented in the TV series in various configurations. Most often, we deal with scenes with a single woman (26% of recording time). Two more configurations, which occurred more or less equally, are scenes with the man alone (21%) and scenes with a mixed dyad, i.e., two people of different sexes (20%). In general, it can be observed that scenes with one person on the screen (woman or man – 47%) and group scenes (dyads and groups – 48%) take up the same amount of time (Fig. 2). But it is the exposure of heterosexual dyads that begs further interpretation. It suggests a strong concentration on topics associated with relationships. This

may hint at the conclusion that although women dominate the exposure, they are usually set in typically feminine contexts. This suspicion is strengthened when we consider the overall structure of the contexts (frames) which form the background for the majority of the scenes (Fig. 3).

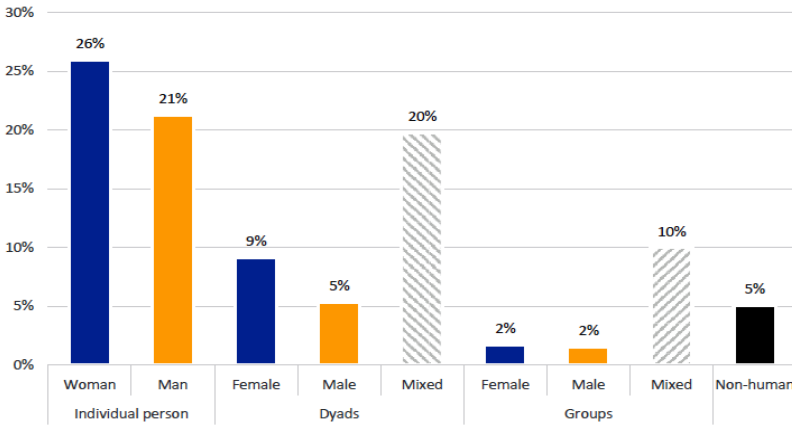


FIGURE 2
Character visibility configurations on the screen.

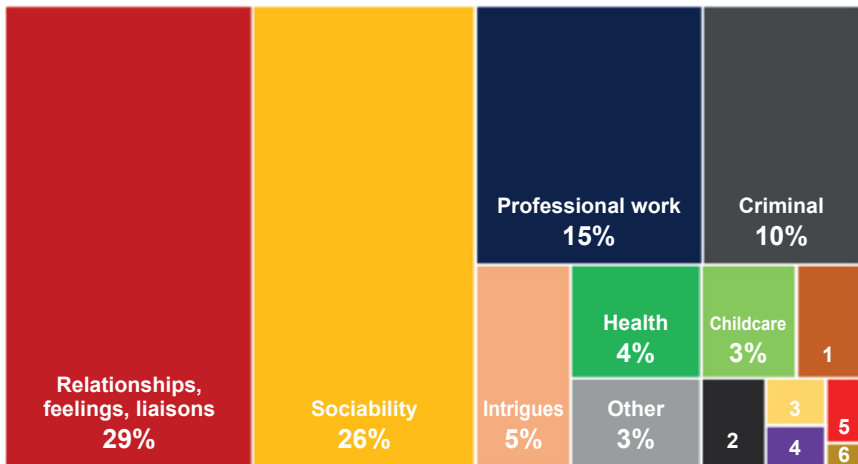


FIGURE 3
Context share in the total recording time.

The plot of the sampled episodes is most often set in the context of relationships and associated emotions (29%). Another type of context in which events happen is sociability (26%), i.e., scenes where ordinary conversations are held, sometimes framed as social events and sometimes reproducing casual daily interactions. The third context in which most of the scenes are embedded is the professional plot (15%), which means in practice that a large number of scenes occur in the workplace of the characters or directly relate to their work. Intrigue and health are less readily found in almost every series. We also recorded plots that, although noticeable in the general analysis, are in practice limited to selected series of a particular genre. These include the criminal context, which appears mainly in two crime dramas and armed conflict, which we see almost exclusively in *War Girls*. The above suggests that scenes connected with emotions, relationships and general sociability constitute more than a half (55%) of the analysed recording time. Had we excluded from the sample two thematic series *War Girls* and the criminal medical drama *Diagnosis*, this domination would have been even greater. Thus, women dominate in stereotypically feminine contexts – cross-reference analysis of active and passive exposure and interaction contexts further strengthens this conclusion.

The proportion of women's active exposure varies from 15% to 53% of the recording time, depending on the context (Fig. 4). Women have the highest share in the statements in scenes concerning childcare (53%), shop-

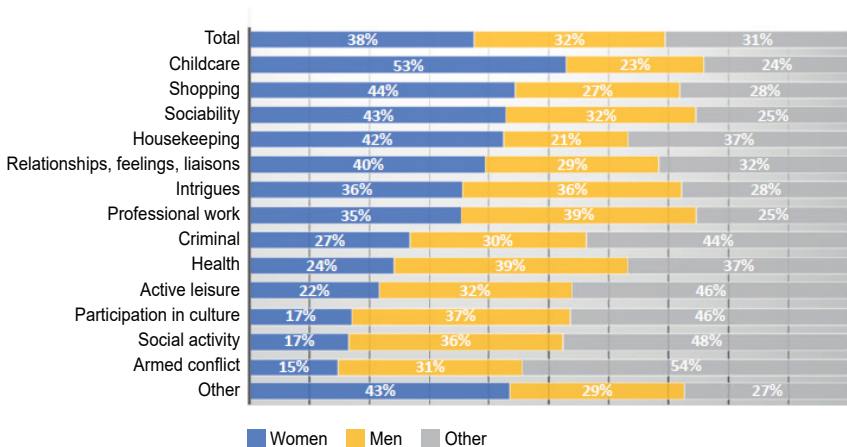


FIGURE 4
Active exposure of men and women in various contexts of interaction.

ping (44%), sociability (43%), housekeeping (42%), and those relating to relationships, feelings, and liaisons (40%). One must remember though that shopping, housekeeping and childcare are the frames for the scenes, constituting roughly only 5.5% of the sample. It is hard to conclude then that women depicted in contemporary Polish TV series are depicted as constrained within the home space. Nevertheless, they seem to be preoccupied with interpersonal and emotional problems constituting most of the scenes in the relationship and sociability contexts. Is it due to culturally embedded stereotypes or the need for TV afternoon formats? Our study does not provide the answer to that question. In contexts related to professional work, health, participation in culture, or armed conflict, men have a greater share of the utterances.

When it comes to passive exposure, we see women appearing more often in the two most popular contexts. Then, rather surprisingly, at work, crime or intrigue, women and men are exposed in a balanced way. In less frequent contexts, women still dominate in the area of housekeeping, children, shopping and active leisure. Men's passive exposure has been long in the context of health, armed conflict, social activities and participation in culture. Such distribution of exposure mostly reproduces stereotypical divisions of gender roles, except for the following plots: criminal (a surprisingly large proportion of women's exposures), intrigues (surprising balance) and active leisure (a slight advantage for women was not obvious) (Fig. 5).

Therefore, it can be concluded that balancing the presentation of gender roles in contexts that may be considered masculine seems somewhat easier than introducing the clear exposure of men in contexts considered closer to feminine. This hypothetical conclusion calls for further studies in the context of research on how genders react to social change in general.

Some further insights into the quantitative results are provided by the qualitative analysis of the gender stereotypes identified in the sample. The analysed series features the most important cultural stereotypes about women. In 4 out of 8 series, we found at least 9 characters depicting the stereotype of the warm homemaker. And although twice this stereotype is overcome in the course of action, it still serves as a convenient scheme to characterise women's engagement with their homes and families. Women in the series are focused on building lasting interpersonal relationships of both an erotic and friendly nature. And often, they occupy the top spot in family structure – in 4 series, we identified matron figures presented quite stereotypically as somewhat older, serious guardians of family hierarchies and orders.

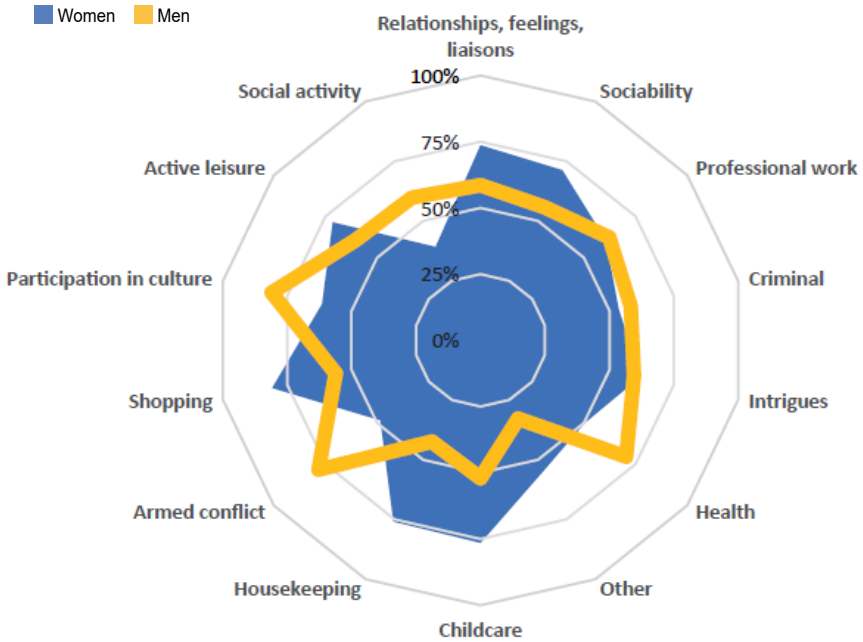


FIGURE 5
Passive exposure of men and women in various contexts of interaction.

This contrasts strongly with the fact that the sampled series rarely presents models of full, unreconstructed, and traditional families. Informal, patchwork and incomplete relationships ‘with a history’ dominate. The a-stereotypical presentation of ‘step-mothers’ is worth mentioning in this context. We found characters defined by their stepmother status in 3 out of 8 series, with all the characters presented as loving mothers who care about their step-children without prejudice and bias, sometimes even better than birth mothers.

More generally, we may conclude that women are shown to be more emotional and sensitive than men. In virtually every analysed series, we found at least one character described by the stereotype of higher feminine emotionality. Sometimes, this emotionality is weaponised by women, as at least 5 instances of stereotypical *femme fatale* suggest. Curiously, these portraits of dark women leading men to their doom seem to bear much attraction for filmmakers, as there are only 3 instances of the somewhat opposite

stereotype of bimbo/blondie. One could conclude then that regarding relationships, the image of women is rather stereotypical. This is however subject to the proviso that if this stereotype were to include women's passiveness, then it is just the opposite. If women feel lonely, they do not hesitate to look for new partners or take the initiative. Sometimes, they try to break up other people's relationships or take revenge on their former partners.

This is better understood when we consider that at the same time and regardless of the concentration on home and relationships, women are shown as professionally active. They are shown in a variety of professional roles, both those traditionally assigned to women and those previously reserved for men. Worth mentioning here are examples of analysed criminal series, in both of which women are depicted as perpetrators, rather than victims of crime. They mastermind the plots and manipulate other people – men and women – in equal measure. This image is surely ambivalent – on the one hand, it undermines the stereotype of a woman as the victim of crime; on the other, it results in female protagonists being portrayed as black characters in both analysed criminal series. The female criminals can hardly be described as passive and this aligns with the general image of women in the analysed sample. When facing danger, they show ingenuity and courage. Women can count on other women – friends or neighbours – in their endeavours and adventures.

Usually, work does not interfere with women's care for their homes and families, because they are, unlike men, very good at multitasking. However, if they become overambitious, their private lives suffer. In as many as 5 series, we registered characters in line with the stereotype of a businesswoman with a miserable personal life. What is interesting is that not only family life suffered for those characters, but other social relations were also equally damaged. Hence, often they look for consolation and escape, for example, through alcohol.

Conclusions

On average, in all the analysed series, female characters predominate both in active exposure and on-screen presence. This difference, especially in the first case, is significant. As the research material contains four genres, there are differences between and within the genres (this applies, for example, to soap operas). The proportions are reversed only for historical dramas.

The series are dominated by contexts related to interpersonal relation-

ships – partner, emotional, or sexual (29%) and relationships of a social nature, such as meetings, visits, or events (26%). Both take over half of the time (55%). Two more contexts are professional work (15%) and crime (10%), but they are very differentiated between individual genres and titles. Contexts that occupy a small percentage of the time are childcare (3%) and housekeeping (2%), while issues such as participation in culture, active leisure, or social activities appear in trace amounts.

The presentation of female characters is two-fold. Firstly, women maintain a stereotypical position as the guardians of the hearth and home and custodians of family relations. Secondly and simultaneously, they do increasingly cross the borders of the male world in the professional field. In the context of the initial research question, it can therefore be argued that the series under analysis confirms the existing gender stereotypes more than it attempts to change them. Where stereotypes are reversed or broken in the course of the plot, it seems more like following the trends observed in contemporary societies, than trying to create a new social reality in the media.

BIBLIOGRAPHY

- Boksański, Zbigniew (1997). *Stereotypy a kultura*. Wrocław: Wydawnictwo Leopoldinum.
- Chłosta-Zielonka, Joanna and Iza Matusiak-Kempa (2015). "Sposoby deprecjonowania sprawczyń wypadków samochodowych." In *Idee i wartości w języku i kulturze*, edited by Iza Matusiak-Kempa and Alina Naruszewicz-Duchlińska, 116-25. Olsztyn: Instytut Filologii Polskiej Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.
- Czernecka, Julita (2011). *Wielkomięscy single*. Warszawa: Poltext.
- Domański, Henryk (1995). "Równouprawnienie. Stereotyp tradycyjnego podziału ról." In *Co to znaczy być kobietą w Polsce*, edited by Anna Titkow and Henryk Domański, 65-87. Warszawa: Wydawnictwo IFiS PAN.
- Gajlewicz-Korab, Katarzyna and Katarzyna Konarska (2017). *Koncentracja kapitału w mediach i jej zapobieganie we Francji oraz Niemczech*. Instytut Staszica. http://instytutstaszica.org/wp-content/uploads/2017/03/IS_koncentracja_mediow_i_jej_zapobieganie_we_Francji_i_Niemczech.pdf (last accessed 14-12-22).
- Garfinkel, Harold (1967). *Studies in Ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Goffman, Erving (1975). *Frame Analysis: An Essay on the Organization of Experience*. Harmondsworth: Penguin Books.
- Hallin, Daniel C. and Paolo Mancini (2012). "Introduction." In *Comparing Media Systems: Beyond the Western World*, edited by Daniel C. Hallin and Paolo Mancini, 1-8. New York: Cambridge University Press.
- Jakubiec, Sylwia (2007). "Macocha, ojczym: stereotypy baśniowe i społeczne a wizerunek współczesny." In *Rodzina – w świetle zagrożeń realizacji dotychczasowych funkcji. Szkice monograficzne*, edited by Wiesława Korzeniowska, Andrzej Murzyn, and Urszula Szuścik, 93-100. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Kochan-Wójcik, Marta (2003). "Problematyka badań nad psychologicznym wizerunkiem kobiecego ciała." *Czasopismo Psychologiczne [Psychological Journal]* 2(9): 159-67.
- Łódzki, Bartłomiej (2008). "Establishing the Dual Television System in Poland." In *Comparing Media Systems in Central Europe: Between Commercialization and Politicization*, edited by Bogusława Dobek-Ostrowska and Michał Głowacki, 99-110. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Mandal, Eugenia (2004). *Podmiotowe i interpersonalne konsekwencje stereotypów związanych z płcią*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Radziejewski, Bartłomiej (2018). "Polska racja stanu w srebrnym wieku", *Nowa*

- Konfederacja* 4. <https://nowakonfederacja.pl/polska-racja-stanu-w-srebrnym-wieku/> (last accessed 14-12-22).
- Simmel, Georg (2009). *Sociology: Inquiries into the Construction of Social Forms*. Vol. 1. Translated and edited by Anthony J. Blasi, Anton K. Jacobs, and Mathew Kanjirathinkal. Leiden-Boston: Brill.
- Sztompka, Piotr (2020). *Słownik socjologiczny: 1000 pojęć*. Kraków: Znak Horyzont.
- tw (2018). "TVP straciła 5 proc. widowni, ale dalej wyprzedza Polsat ze stacjami ZPR i TVN z Discovery. Szczyt oglądalności na 'Faktach'", 10 April. <http://www.wirtualnemedial.pl/artykul/tvp-stracila-5-proc-widowni-ale-dalej-wyprzedza-polsat-ze-stacjami-zpr-i-tvn-z-discovery-szczyt-ogladalnosci-na-faktach> (last accessed 14-12-22).
- Zybertowicz, Andrzej (1995). *Przemoc i poznanie*. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.

LIST OF THE SERIES

- L for Love* [*M jak Miłość*] (2018/2019)
- First Love* [*Pierwsza Miłość*] (2018/2019)
- Crown of Kings* [*Korona Królów*] (2018/2019)
- War Girls* [*Wojenne Dziewczyny*] (2018)
- Diagnosis* [*Diagnoza*] (2017)
- Trap* [*Pułapka*] (2018)
- Girls from Lviv* [*Dziewczyny ze Lwowa*] (2018)
- Friends* [*Przyjaciółki*] (2019)

WOMEN IN POLISH TV SERIES: QUALITATIVE AND QUANTITATIVE ANALYSIS



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.034c67d8>.

In Giorgio Avezù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

7. Queer and Mainstream: Tracing *Heartstopper's* Reception via Digital Tools

Marta Boni

◀ ABSTRACT

Netflix's *Heartstopper* (2022) raises a paradox for queer television: does the series present, radically, a queer utopia, or does it merely serve commercial imperatives? Exploring this phenomenon with the help of digital tools, I will put forward a methodological discussion centered on the need of uncertainty as a core feature of the contemporary structure of feeling, as well as a crucial caveat for any foray into the analysis of television.

KEYWORDS

TV studies; digital humanities; LGBTQIA+; queer.

Introduction

This paper is part of a set of considerations in the field of digital humanities. It stems from the belief that the alliance between media studies and digital humanities is more necessary than ever in a context where media production and consumption are done through computer screens, and in a networked landscape. It is now easier than ever to be confronted with mass amounts of data. However, it is imperative to make room for emotional elements, which are always present in these aggregates of data. The question is: can a data-driven perspective offer us a new way of reading television phenomena and, more broadly, the audiovisual?

We can justify the interest in using data-driven perspectives for television analysis, given the very nature of the televisual medium itself. This is true for at least two reasons: first, a TV program is often a serial object, comprised of a plurality of episodes and, oftentimes, seasons. Second, the contemporary landscape produces an impressive amount of content, making it impossible to understand local or general trends without some strategy for mapping it.

Some scholars in media studies apply data-driven perspectives, often placing them within the discussion of Franco Moretti's distant reading, a project of studying literature "from a distance" by compiling large quantities of data. Unlike the traditional close reading, apt for analyzing large corpora (Moretti 2005, 2013), such an all-encompassing approach simplifies certain analytical duties. For example, counting discrete elements of a film or a show, like Yuri Tsivian's *Cinematics* or Jeremy Butler's *Shot Logger* ("Cinematics n.d., Butler [2014] 2020), or visualizing them differently (the *Distant Viewing Project* by Taylor Arnold and Lauren Tilton and their use of algorithms for the analysis of face detection in American sitcoms (Tilton 2019); examples of network analysis can be found in the *Narrative Ecosystems* projects (Pescatore et al. 2014, De Pascalis 2015, Innocenti and Pescatore 2017).

These approaches revolve around formal or narrative features of TV shows. They are informed by a theorization of the series as structures, in which links and patterns reveal specific, meaningful features. Such “distant reading” or “distant viewing” practices are grounded in structuralist rigor and a certain optimism, or even positivism, about the idea of seeing better from afar, thanks to the use of big data. Taking a step back as a way of seeing differently, of seeing better, rests on a certitude: the exactitude of numbers leads to exportable results, which should be more “objective” and less biased than a single set of close analyses or idiosyncratically constituted corpora. Taking Moretti’s position as a starting point for new methods in the analysis of film and television may seem like a way of trading the impossible task of analyzing every element of a long, perhaps never-ending TV show, with the option of obtaining a clear, broad view of it. Yet, even if such a perspective was presented by Moretti as groundbreaking, such a view does not disrupt the norm. It is a “strong” strategy, constructed in clearly defined settings, using specific software.

Digital cartography and data visualization can benefit from the “quantitative condition” of our times but, as I argue, it should be done knowing there will be an inevitable amount of uncertainty. Therefore, a queer approach enables a critical view towards our methods. This is particularly true because television shows are phenomena composed of multiple, transmedial, and often anonymous contributions, which go beyond the boundaries of official textuality. Fan discourse, for example, creates a tangle of official texts, paratexts and grassroots readings, prompting a need to adopt more “uncertain” queer practices, able to display such disorder, but without necessarily trying to organize it. We can therefore ask: “what might it mean to do the work of the digital humanities queerly?” (Ruberg, Boyd, Howe 2019, chap. 8, page 2)

When dealing with the queer, a digital humanities perspective is particularly useful. It allows intimate discourses to emerge, it foregrounds conversations that only take place online due to the delicate nature of the subject matter, and helps us display fragments of discourses that would otherwise remain inaccessible. Doing queer research in digital humanities can mean tackling queer issues, and building and dismantling tactics to adapt them to queer sensibilities. Television is a particularly appropriate terrain for such work. Traditionally, it is a space where social values are discussed. According to John Fiske, television is a

[b]earer/provoker of meaning and pleasures, and of culture as the generation and circulation of this variety of meanings and pleasures within society.

Television-as-culture is a crucial part of the social dynamics by which the social structure maintains itself in a constant process of production and reproduction: meanings, popular pleasures, and their circulation are therefore part and parcel of this social structure (Fiske 1994:1).

Thus, as we consume online, we leave fragments, whether it is feeding the algorithm of the video-on-demand platforms, liking content, sharing an image or leaving a comment. Other fragments include the heterogeneous operations of fans, from the manufacturing of visual art to fanfiction, fanvid, cosplay, etc. I will suggest, while developing strategies based on quantitative tools for assessing discourse analysis, that our answers need to address the issue of uncertainty contained in an abundance of online traces, yet without trying to fix it. Moreover, our answers may promote social justice by bringing into view the entanglement of forms of oppression. It is therefore particularly important for anyone who deals with these practices to commit deeply to the studied sensibilities.

Far from wanting to merely expose them, the aim is to let traces of discourse appear that can be helpful in highlighting a show's nuances. As we will see, we can highlight the relevance of queer issues in mainstream media, such as the presence of what we will call "queer melancholia." In applying a queer approach, which embraces imperfection, the limits of our knowledge, and the "weakness" of our practices, what perspectives and methodologies are possible in digital humanities?

A TV series is built up from its official texts, but also from its "digital traces" (Merzeau 2012). My perspective, rooted in cultural studies, aims to highlight the relevance of what I call "traces of reception" (Boni 2017). Audiences discuss television shows online and use their life experience to make sense of a series, and vice versa (they make sense of their lives by using TV texts as indicators, guides, or environments to explore). Like tracks along a path, such data can lead to a hypothesis and findings, but they probably won't be sufficient to describe the entire phenomenon. They will only tell us where *we are* within the phenomenon itself. And therefore, such attempts lay bare the impossibility of an orderly typology, of a precise track to follow. Embracing a queer episteme (keeping in mind that queer practices center on deviations from the norm, and the questioning of the norm itself), I find myself happily working with a fuzzy infrastructure. And I develop a queer praxis, assembled with recursive explorations of a territory and from a series of attempts.

The heuristic concept of emergence, found in Morin's work on complex systems, is particularly relevant here (Morin 2005). The idea that disruptions of a system can appear and lead to unforeseen results is compatible with an approach that questions the norm. Yet, questioning the norm is not possible unless we scrutinize the methods themselves. It is my aim to discuss further the use of data-driven approaches for the analysis of television shows. The chosen example points towards a bricolage of data-gathering techniques, and to the relative success of digital tools that unveil one singular feature of the observed phenomenon, which, far from offering a clear answer, leads to a more uncertain one, an answer "in-between."

And to find a path in the "messiness" or "playfulness" of queer digital practices, instead of the cleanliness and rigor of traditional work, I am suggesting a mixed methodology, a perspective that uses both close reading and its potential to reveal poignant details in the show, and the potential of quantitative analysis to reveal patterns, to adopt a queer perspective. In other words, analytical tools that make room for disruptive, non-normative, unlabeled feelings to appear on the map.

The logic of television – the fact that it is a mass medium, crossed by incessant reformulations, mutations, stops, failures, and intervals – also justifies the use of a queer approach. It generates a need to borrow from queer theory its particularly unique instruments that can help us better understand television. TV intended as a "cultural forum" (Newcomb and Hirsch 1983) has always been a space for resistant sensibilities, just think of the queer fandom highlighted by a consistent literature since the '90s (Jenkins 1992, Stein 2015, Lothian 2018). So, beyond the "negative" perspective that the queer rift can bring, we can envision, with José Esteban Muñoz, a "utopia," a possibility of reworking the present from a queer "not yet" envisaged in the future, evoking surprising and unexpected alternatives (Muñoz 2009). As we will see, when we enter such a heterogeneous space, uncertainty, disorientation, and even failure are at the core of our methodological discussions.

Heartstopper

While LGBTQIA+ characters have always had a rather uneasy life on the small screen (think of the cliché of gay characters condemned to secondary parts or to the role of the villain, destined to suffer and die, or to disappear because of the series' cancellation), today there is a greater, more nuanced

presence of minorities in serial productions. Since *Orange is the New Black* (Netflix, 2013-2019), digital platforms and, more timidly, linear television, produce stories featuring more overtly LGBTQIA+ characters, responding to a need for representation expressed by an increasingly visible audience (and creators). Netflix's algorithms also contain information on diversity, with title descriptions incorporating "non-normative identities," and the company is said to be actively working to be more inclusive both behind and in front of the camera. The renewal of content goes hand in hand with access to production, writing and directing that is slowly becoming more equitable for women, racialized people, and members of the LGBTQIA+ community. Gradually, the reconfiguration of norms and the appropriation of issues of "identity" are in turn becoming a standard in the creation of scripts that are increasingly labeled as "queer."

Netflix's *Heartstopper* (2022) is an adaptation of Alice Oseman's light and romantic graphic novel for young adults, "where just a blush or a smiley-face emoji convey so much more than words" (Martaus 2020). The story focuses on the budding romance between two young boys. Charlie and Nick are teenage friends who attend a British school. Charlie, a bit isolated and nerdy, has been openly gay some time, whereas Nick, athletic and popular, doesn't know how to label his orientation, or hasn't questioned it yet. The series follows, over eight episodes, the stages of a discovery: the crush, the development of their love, and the moments when their hearts crack or stop for a moment. The presence of homophobic judgment or violence from classmates does not prevent the show from being optimistic.

The episodes follow the wonder, doubt, and enthusiasm of Charlie and Nick, and offer glimpses into the lives of secondary characters who also experience moments of transformation or who face their feelings for another person. At the core of the show, vulnerability and even benevolence – elements still underestimated in prestige serial productions – take a leading role. It appears through a staging of ordinary situations, touching because seen through a proximity to the two protagonists. We find it in Charlie and Nick's hesitations and clumsiness, as well as in their reciprocal interactions and in their relationships with family and friends. The adult world, on the other hand, remains free of the judgment and conflict that other teen series typically reserve for it. Charlie's father, for example, appears only when he picks up his son at night after a party or movie screening, and in those rare moments he conveys a great gentleness, carried mostly by apprehension for the bullying Charlie may experience. Nick's mother (Olivia Colman, who

wept with emotion at the set during the final sequence) also remains in the background of the action, but plays a reassuring, welcoming role.

Heartstopper poignantly – and minimally – dramatizes some classic moments of a queer adolescence: falling in love with a straight person, the fear of judgment from others, the pain of dating someone who is ashamed of the relationship, homophobic bullying, the deep silence that precedes every coming out. The style attempts to correspond with the original comic book, in that it insists not on the virtuosity of the camera, but rather on drawn elements inserted into the image, conveying the characters' feelings in an almost literal or playful way. For example, when the two boys have butterflies in their stomachs, we see small butterflies appear around them. Leaves, sparks, or hearts drawn with pastel colors appear in the shot, adding a surplus of emotions, facilitating the approach to the characters' sensibilities.

The series does not propose the well-known stereotypes about gay characters who suffer through uncomfortable moments or painfully come out of the closet. Instead, it establishes the elements for a happy ending, offering a fair, optimistic, and non-pedagogical view of the relationship between people of the same sex and their experiences as young folks:

Heartstopper is the anti-*Euphoria* [...] They are sweet, chronically polite and partial to an early night – a world away from the contemporary teen TV stereotype. [...] a comfortingly escapist, binge-worthy beacon of loveliness and love (Aroesti 2022).

It makes the positive parts feel real, and it shows that they *can* be real (Wickman 2022).

And yet, such cheerfulness (which never interjects in sexual moments) could be easily taken for an oversimplification. If it appears devoid of any radical claim, it is obviously a positivist representation:

It's curious that in an age of unprecedented visibility for LGBTQ communities, the queer teens chosen for the cinematic spotlight appear so allergic to, well, seeming gay... Whether the implied assimilationist impulse reflects the filmmakers' or the characters' caution is up for debate. Either way, coming-out stories tend to be told with straight audiences in mind (Li 2022).

The Paradox of Queer Television

While an openness to new themes is somewhat satisfying for marginalized communities, it is double-edged. This is part of the paradox, as Lynn Joyrich and Julia Himberg underlined in a seminal contribution (Joyrich 2014). Being queer means, among other things, taking a step back from the “normative productivity” typical of industrial contexts. Namely, it refers to the refusal of an idea of the future as a space of reproduction (Edelman 2005). The approach is incompatible with the expected progression of a series, a commercial product, within an established media. Indeed, the queer corresponds to the “subversion of the ordinary, the strange, the irregular, what would seem to require a disruption of some kind of our ‘regular programming’ (Joyrich 2014: 134).” The fact that the depiction of queer characters is becoming more and more frequent represents, for Joyrich and Himberg, a confirmation of “negativity” raised in queer theory. The “mainstream” does its best to absorb any possible queer disturbance by suffocating it under its business strategies. Striving for integration leads to a softening of everything radical or rebellious.

Within a landscape where inclusiveness matches market imperatives, television producers aim to reduce uncertainty, to control the unknown. Consequently, this creates a binary separation between the reason and the wild (Halberstam 2020), even when addressing troubling characters or features. Indeed, representing LGBTQIA+ characters on television does not automatically equate to the expression of a queer sensibility. Moreover, it is not necessary for oblique and queer paths to represent characters belonging to this community: queerness can be found in shows that display heterosexual characters looking for an alternative to the dominant patriarchal culture. As a matter of fact, it can boost the establishment of new forms of homogeneity. As Jack Halberstam well says, “queerness without wildness is just white homosexual desire out of the closet and in synchrony with a new normal” (Halberstam 2020: 39). This is also the starting point of Joëlle Rouleau, who states... “that signifier [LGBTQ+] becomes a placeholder for neoliberal, value-hungry strategies to reify the inclusion of sexual diversity, in particular in media representations” (Rouleau 2021).

By analyzing the reception of this series, we will get closer to understanding the nuances of this paradox. At the same time we can ask a methodological question about the fragile, fiddled nature of digital humanities, as well as the political issues linked to extracting, mapping, and visualizing data. Can our tools be queer?

Queering our Methods

Viewers streamed it for almost 24 million hours in the week between April 25 and May 1 (“Heartstopper News on Twitter” 2022). *Heartstopper* is “100% fresh” according to the aggregator Rotten Tomatoes. We can see a relevant consensus in the main anglophone and occidental spaces. In order to gather more details about this consensus, I analyzed the online reception of the series as it emerges from a web scraping of the microblogging platform Twitter and Reddit, the social news aggregator and discussion website.

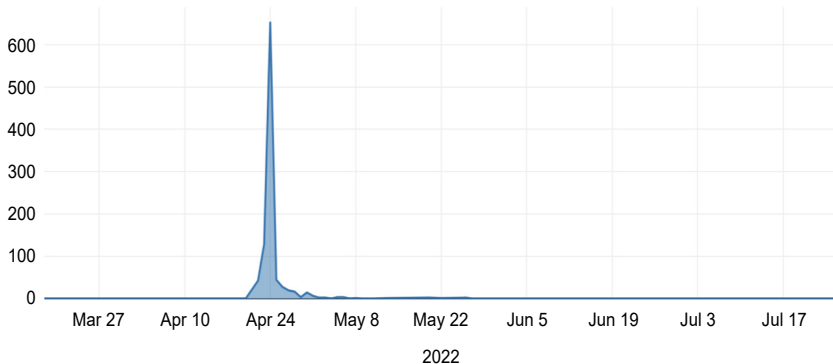


FIGURE 1
Twitter conversation about *Heartstopper* around the release date.

Each data pulling session allowed me to extract 1,000 tweets (due to the constraints of the account – Pro account on Commanalytic, trial version). I carried out 3 extractions: 2 on Reddit (over three consecutive days, twice), one on Twitter. Plus, as we will see, I explored the francophone space through a tool developed by the Institut National de l’Audiovisuel. These data allow me to paint a picture (a limited, biased one) of *Heartstopper*’s reception, or one of its possible readings. This is to be understood within a larger observation to define a few constants, a few “objective pieces of evidence”: the fact that the show “functioned”, for example.

Reddit is an online community where registered users share links, images, videos, and comments about any cultural item. Contributions are organized by subject into user-created boards called “communities” or “subreddits”, and each is voted “up” or “down” by the community. Such a space functions as an intermediary between the viewers’ experience and the series as a collective object.

To extract data, I used the Commanalytic EDU Subreddit Historical Data Collector, which collects available posts (including submissions, comments and replies to comments) from a public subreddit for up to 7 consecutive days. Then I ran some of the results through Voyant, an open-source, web-based application for text analysis. Such tools are open access, and do not require coding expertise. They are the result of a bricolage, driven by the need to follow a show's digital traces. It should be noted, as mentioned before, that my exploration follows a more recursive path rather than a linear one. I alternate between different sources and platforms, and collect results over multiple sessions, then use historical data to pick specific moments of the show's life to observe their relevance.

I already pointed out how pertinent it is to consider visualizations and digital cartography as heuristic practices (Boni 2018), and to stress the constructed nature of our findings. The use of data-driven perspectives does not replace the subjectivity of close analysis. Indeed, the material we gather, even when considering big data, changes shape according to the point of view used. According to Johanna Drucker, information is not naturally present in any online space (Drucker 2014). Instead, it is always *capta*; in other words, the result of a gathering process originating from the application of a hypothesis, or a research question. It is the product of how information is organized depending on the platforms we use. Far from being a structured environment, a series is a scattered, ever-expanding space whose scope and properties depend on the adopted perspective as well as the tools selected. Our observations are dependant on the tools we use, too. Most of the time, a quantitative web analysis / web ethnography relies much more on the shape of the platform than on a real phenomenon (Taurino and Boni 2018).

For example, on Twitter, we notice a spike in conversations about the show two days after its release. Unsurprisingly, this shows the relevance of binge-watching practices and gives a sense of just how the release of a show can be seen as an event. This is particularly true in the case of a series which already had a fan base due to it existing in another medium. For *Heartstopper*, it was the graphic novel. In the comments, we can observe the merging of the two media. The spike in comments around the release date also means that viewers could watch the whole show in a relatively short period of time. This also stresses the effectiveness of short series with short episodes (a contemporary trend) employing an efficient marketing strategy focused on the release. There is a precipitous drop in density after the second week. It is also apparent that we are facing a short-lived attention to the series; the show is

not a “big” phenomenon in the media sphere, despite an enthusiastic ongoing discussion about its success in portraying queer characters.

Still, when we compare such a limited time frame to the French INA web archive tool WebMédia we see the debate about HS does not stop at all. Fan chatter persists, making many different links to other media, especially the graphic novel.

Quantitative analysis of the phenomenon also facilitates, on Twitter, the extraction of frequently used terms (word cloud) or emojis, which is specific to Twitter conversations but also to *Heartstopper*'s target demographic (the use of emojis directly inspired from the show's graphics).

Network analysis, made possible also via Commanalytic, allows us to inquire about the shape of the conversation. We can focus on the link between some accounts and the replies they get. Alice Oseman, the author, is placed at the center of the network. When we visualize the amount and density of nodes, of interactions, many small, disconnected interactions appear.

When we use the sentiment analysis tool, something interesting emerges. The vast majority is positive, another big chunk neutral, while only some posts are negative. And that's where the element of uncertainty emerges: the impossibility to measure, only by quantitative data, one of the show's specificities. What could be negative within a sphere of discourse that, as the media relate, is homogeneously satisfied with the show?

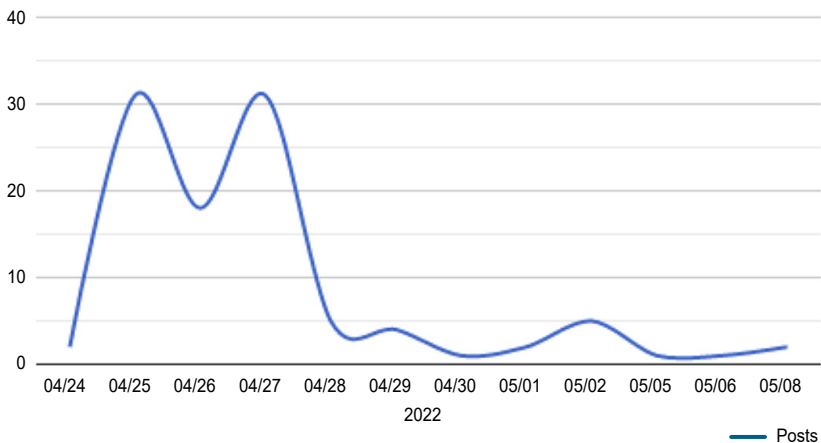


FIGURE 2
Reddit posts over time (via Netlytic).

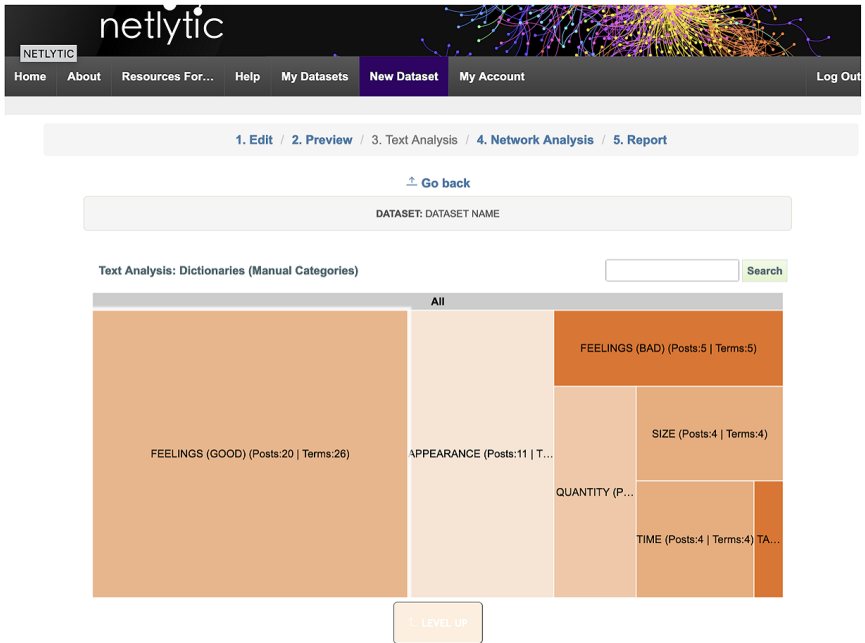


FIGURE 3
Reddit: sentiment analysis (via Netlytic).

Allowing space for technical bias, we can still look for this negativity, treating it as one of the *emergences* of the system, something that we can observe better from afar. What could be negative about HS? Without a precise term or clue, our corpus does not tell it yet. We must analyze another corpus, and progress deeper in our experience of the fan community to find some evidence of this “negativity.” Let’s consider this comment: “Seeing Nick and Charlie’s story on screen was saddening because I never got to be in love in my teenage years and I wish I could’ve been (Cleal n.d.)”

This raises the relevance of a “queer negativity” within the overall extremely positive reception. We think of Lee Edelman and his claim about the impossibility for queers to look forward to a future, and towards the reproducibility typical of heteronormative, patriarchal routes. Such queer negativity is also explored by Love, when she talks about queer backwardness, or the feeling of melancholy which accompanies queer people having suffered in the past and who still feel it’s impossible to “fit,” or satisfy their desires (Love 2009).

So, on Reddit, I found, as would be expected, several enthusiastic posts about the show on subreddit *0/heartstopper*. Yet, upon closer inspection, there was a significant number of posts focused on a very peculiar feeling: a mix of melancholy, nostalgia, and mourning. Many users expressed their admiration for the show and the pleasure they derived from its happy ending and positive relationships, a situation that, during their teen years, they never experienced. As queers, or even as straight people who declare themselves currently in a satisfying relationship, they find themselves grieving their past. Rewatching the series many times, or reading the graphic novel is the solution they find to get over what the community now calls the “HS syndrome.”

Watching these grinning teens throw themselves into love with abandon prompted a twinge of melancholy. Was this even a possibility for me at that age? Would I have jumped at this opportunity like they do or was I too tied up in self-conscious knots to let myself leap?
It's hard to be a human in the year 2022, and so we all need our little treats.¹

Conclusions

This dive into digital tools has shown the potential for developing a queer praxis for studying a series/media content as a space of discourse. The tools used are open access, do not require coding skills, and allow us to display traits of a series' reception and its circulation. Over the years I've noticed that using an approach tailored to the phenomenon we want to observe gives us a more holistic view of the cultural dimension of the media, and their status as a commodity in the current economy. Quantitative analysis, matched with qualitative, highlights several interesting features that a close reading wouldn't reveal: a conversation's time frame and therefore its link to specific moments of a show's life; the frequency, density and variability of such discourse; the most used keywords; and networks of relations. Of course, such a method is fragmentary, or even rudimentary. Many adjustments could be made. We have also seen the limits of using third part apps. I also suggested the possibility of comparing results of different tools without any expectation of finding one that would be better than the others. More than anything, the outcome is a methodology based on bricolage, a

¹ <https://www.reddit.com/user/Next-Ordinary-2491>

relativism that is not a renunciation of scientific rigor, but instead, it is the acknowledgment of the situated nature of each observation. Digital humanities will remain human precisely because it welcomes trial and error. We keep testing tactics (as De Certeau would say: the “strategies” of the weak (Certeau 1980)) and adapting them to our objects, surrendering the ideal of recreating the “real” image of the phenomenon. But maybe, by following some traces we find across the path, we will better understand our relationship to our objects, and raise different questions, embracing uncertainty, crisis. And choosing to answer them without a clear, defined, strong answer.

BIBLIOGRAPHY

- Arnold, Taylon and Lauren Tilton (2019). "Distant viewing: analyzing large visual corpora." *Digital Scholarship in the Humanities* 34(Suppl 1): i3-i16. <https://doi.org/10.1093/llc/fqz013>.
- Aroesti, Rachel (2022). "No Drugs, Booze, Sex or Swearing: Will *Heartstopper* Rewrite the Young Love Rulebook?" *The Guardian*, March 18. <https://www.theguardian.com/tv-and-radio/2022/mar/18/no-drugs-booze-sex-or-swearing-will-heartstopper-rewrite-the-young-love-rulebook> (last accessed 25-09-22).
- Boni, Marta (2017). "Des téléspectateurs géolocalisés." *Sens Public*, December. <http://sens-public.org/article1283.html>.
- Boni, Marta (2018). "Digital Maps and Fan Discourse: Moving between Heuristics and Interpretation." <https://doi.org/10.25969/MEDIAREP/3464>.
- Butler, Jeremy (2014) 2020. "Shot Logger." PHP. <https://github.com/jgbutler/Shot-Logger>.
- Certeau, Michel de. (1980). *Arts de faire. Invention du quotidien*. Paris: Union générale d'éditions.
- Cinematics (n.d.). "Cinematics - About. <http://www.cinematics.lv/> (last accessed 20-03-23).
- Cleal, Sam (n.d.). "Here's Why I Found *Heartstopper* Equally Traumatic and Gratifying at the Same Time." BuzzFeed. https://www.buzzfeed.com/sam_cleal/heartstopper-review (last accessed 25-09-22).
- De Pascalis, Ilaria A. (2015). "The Gotham Ecosystem: Characters Networks and Visual Influences in the Batman Franchise." Presented at the *Cinema/Series. A Two-Day International Workshop*, Universitat Politècnica de València, November.
- Drucker, Johanna (2014). *Graphesis: Visual Forms of Knowledge Production*. MetaLABprojects. Cambridge: Harvard University Press.
- Edelman, Lee (2005). *No Future: Queer Theory and the Death Drive*. Durham: Duke University Press.
- Fiske, John (1994). *Television Culture*. London: Routledge. <https://doi.org/10.4324/9780203837153>.
- Halberstam, Jack (2020). *Wild Things: The Disorder of Desire*. Durham: Duke University Press. "Heartstopper News 🐦 sur Twitter." 2022. Twitter. May 3, 2022. https://twitter.com/Heartstopper_TV/status/1521566181257269248.
- Innocenti, Veronica and Guglielmo Pescatore (2017). "Narrative Ecosystems: A Multidisciplinary Approach to Media Worlds." In *World Building. Transmedia, Fans, Industries*, edited by Marta Boni, 164-83. Amsterdam: Amsterdam University Press.
- Jenkins, Henry (1992). *Textual Poachers: Television Fans & Participatory Culture*. New York: Routledge.
- Joyrich, Lynne (2014). "Queer Television Studies: Currents, Flows, and (Main)Streams." *Cinema Journal* 53(2): 133-39. <https://doi.org/10.1353/cj.2014.0015>.

- Li, Shirley (2022). “‘Heartstopper’ and the Era of Feel-Good, Queer-Teen Romances.” *The Atlantic*, May 11 <https://www.theatlantic.com/culture/archive/2022/05/heartstopper-netflix-love-victor-coming-out/629827/> (last accessed 25-09-22).
- Losh, Elizabeth and Jacqueline Wernimont (2019). *Bodies of Information: Intersectional Feminism and the Digital Humanities*. Minneapolis: University of Minnesota Press.
- Lothian, Alexis (2018). *Old Futures: Speculative Fiction and Queer Possibility*. New York: New York University Press.
- Love, Heather (2009). *Feeling Backward: Loss and the Politics of Queer History*. Cambridge: Harvard University Press.
- Martaus, Alaine (2020). “Heartstopper: Volume 1; by Alice Oseman (Review).” *Bulletin of the Center for Children’s Books* 73(10): 442-43. <https://doi.org/10.1353/bcc.2020.0437>.
- Merzeau, Louise (2012). “Faire mémoire de nos traces numériques.” *E-dossiers de l’audiovisuel*, June, 33840 signes (article en ligne).
- Moretti, Franco (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- Moretti, Franco (2013). *Distant Reading*. London: Verso.
- Morin, Edgar (2005). *Introduction à la pensée complexe*. Paris: Seuil.
- Muñoz, José Esteban (2009). *Cruising Utopia: The Then and There of Queer Futurity*. New York: New York University Press.
- Newcomb, Horace M. and Paul M. Hirsch (1983). “Television as a Cultural Forum: Implications for Research.” *Quarterly Review of Film Studies* 8(3): 45-55. <https://doi.org/10.1080/10509208309361170>.
- Pescatore, Guglielmo, Veronica Innocenti and Paola Brembilla (2014). “Selection and Evolution in Narrative Ecosystems: A Theoretical Framework for Narrative Prediction.” In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 16. IEEE. <https://doi.org/10.1109/ICMEW.2014.6890658>.
- Rouleau, Joëlle (2021). “Goopy Connections: A Little Detour En Route.” *Jump Cut*. <http://www.ejumpcut.org/archive/jc60.2021/Rouleau-QueerTVsection/index.html>.
- Stein, Louisa Ellen (2015). *Millennial Fandom: Television Audiences in the Transmedia Age*. University of Iowa Press.
- Taurino, Giulia and Marta Boni (2018) “Maps, Distant Reading and the Internet Movie Database: New Approaches for the Analysis of Large-Scale Datasets in Television Studies.” *VIEW Journal of European Television History and Culture* 7(14): 24-37. <https://doi.org/10.18146/2213-0969.2018.jethc151>.
- Wickman, Kase (2022). “Why Has Everyone Fallen in Love With Netflix’s ‘Heartstopper’?” *Vanity Fair*, May 6. <https://www.vanityfair.com/hollywood/2022/05/heartstopper-netflix-alice-oseman-interview> (last accessed 25-09-22).

QUEER AND MAINSTREAM: TRACING HEARTSTOPPER’S RECEPTION VIA DIGITAL TOOLS



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.d0e0bbeb>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

8. TV Reception via Social Media Analysis: The Case of *Doc – Nelle tue mani*

Stefania Antonioni, Dom Holdaway

◀ ABSTRACT

This chapter adopts a data-led approach to studying the reception practices and discourses on social media of a TV series: *Doc – Nelle tue mani*. The series is one of the more traditional flagship products of Italian public broadcaster Rai in recent years, and one of the most successful Italian medical dramas. The essay examines user interactions with the show on two social media platforms – Instagram and Youtube – focusing, in particular, on the prevalent themes in comments and discussions and the kinds of visual content and images remediated from *Doc*. Data regarding user interactions is scraped using a variety of digital tools and then subjected to quantitative and qualitative readings. The analysis indicates that the COVID-19 pandemic, which has affected both the production and the storyline of the show, is a recurrent theme in user interactions on Instagram. It also indicates that the most popular remediated content on YouTube refers to the show's music and to its stars, and first and foremost the lead actor Luca Argentero.

KEYWORDS

TV reception; medical drama; *Doc – Nelle tue mani*; Instagram; YouTube.

Introduction

This paper focuses on the digital “lives” of contemporary Italian medical dramas. Specifically, it seeks to examine (i) the reception of this TV genre on social networks, looking at the discourses of fans and viewers in social media comments; and (ii) the terms in which content from these shows (images, scenes, music) is reproduced online. The research findings presented here¹ are a part of the Urbino unit’s work on the Italian State-funded PRIN project *Narrative Ecosystem Analysis and Development (NEAD) Framework: A Systemic Approach to Contemporary Serial Products. The Medical Drama Case*.² The project is devoted to a macro-level analysis of the medical drama, which is used as one possible example of the “narrative ecosystem” (Pescatore 2018, Brembilla and De Pascalis 2020), principally over three geographical areas – China, the USA and Italy. The objective of the University of Urbino’s research unit is the Italian medical drama; our research focuses on the intersections of its textual elements with its production, distribution and reception contexts. In line with this objective, in this article we focus in particular on the online reception and activities of social television (cf., e.g., Schröder 2018, Ranjana and Ytre-Arne 2018, Livingstone 2013, Ducheneaut et al. 2008) relating to one particularly relevant case study, in order to question not only the broader tones of its reception, but also the national specificity of those reactions. Indeed, following the central premise of the “narrative ecosystem” model, reception is one of a few contingent factors that can influ-

¹ The authors conceived and realized the research collaboratively. For this essay, the authors collectively wrote the introduction and conclusion, Stefania Antonioni wrote section 1 and Dom Holdaway wrote section 2.

² Further information about the project, whose Principal Investigator is Guglielmo Pescatore (University of Bologna) can be found at the following link: <https://dar.unibo.it/en/research/research-projects/prin-narrative-ecosystem-analysis-and-development-framework-nead-framework-un-approccio-sistemico-al-prodotto-seriale-contemporaneo-il-caso-del-medical-drama> (last accessed 21-12-22).

ence the narrative of a serial product externally, leading writers or producers to change the product's direction in response to the audience. Put in other, complementary terms, to demonstrate this link we could invoke to the notion of the "audience in production" as identified by Elena Levine in her study of the medical drama *General Hospital*, to capture the direct impact that (implicit or real) viewers have on set (cf. Levine 2001: 78-80).

Our approach in the essay shares the *NEAD* project's conviction that the medical drama is a particularly relevant and useful object of study in this regard – especially when it comes to reception, as we will indicate. This is both in due to the genre's relevance within mediascapes (often medical dramas are among the longest and most successful TV products, as in the cases of *ER* and *Grey's Anatomy*), but also due to a specific connection between institutions, media, audiences and public discourse – something that is all the more relevant in the years of the pandemic (cf. Rocchi 2019).

Specifically, here we focus on the most popular and successful Italian medical series of recent years: *Doc – Nelle tue mani* (2020-), aired in prime-time by the first channel of public broadcaster, Rai 1 and co-produced by Lux Vide and Rai Fiction. Inspired by the true story of the doctor Pierdante Piccioni, the series narrates the experience of doctor Andrea Fanti, a severe and unfriendly but talented doctor who loses 12 years of his memory when shot by the grieving father of a patient who died at the hospital (the 'real' Piccioni lost his memory following a car accident). The amnesia takes Fanti back to an earlier version of his own life, before his separation from his wife, the loss of his son, and the hardening of his attitude toward patients. Like many other medical dramas, the series is structured around both vertical and horizontal narratives, between single patient stories that usually last just one episode and the overarching story of Fanti and the other medical staff. The series has proven to be a great success in Italy, attracting significant audience shares and a great deal of public and critical attention. Across both seasons, the episodes reached between 6 and 7 million viewers on average, and around 29% of the share; this included in particular a young audience (cf. Fumarola 2020, Nicholas 2022). Following this success, it has also been broadcast on television and via streaming in other European, North and South American and even Far East markets.

Doc's popularity also makes the show a useful case study from the perspective of reception analysis. Broadcast in the present (the first two seasons were released in 2020 and 2022, the third is in production at time of writing), the show is emerging at a time when social media marketing for

public broadcasting is at an unprecedented peak, both in terms of output and of audience accustomization. As we will see later, the first part of the first season was released during the COVID-19 pandemic – which is also referenced directly in the show’s second season – hence during a moment of more intense online presence and social media use. With this in mind, our aim was to examine social interactions relating to *Doc* and remediated content from the series, with two macro objectives. The first was to understand the prominence of the different medical drama plot “isotopies” identified by Pescatore and Rocchi (2019), i.e. the sentimental, professional and medical. The former two, which refer respectively to the personal-emotional and professional relationships of the characters, are horizontal narratives or running plots that transcend single episodes; the third category refers to the medical “mystery” represented by a (usually) new patient, that is often resolved through a vertical/anthological, single-episode narrative (Pescatore and Rocchi 2019: 110-12). It should be noted, though, that in the case of *Doc*, the medical isotopy also includes the Coronavirus, which was represented in the second season’s narrative of the show in a more horizontal manner (as well as having interrupted the production and distribution of the first season). Tracing the isotopies within the show’s reception, we moreover sought to study which of these trigger greater audience engagement.

Our second objective was to interrogate what kinds of images from the show are predominant on social media, therefore signaling, hypothetically, the terms of the show’s imaginary. We sought to ask which actors/characters, settings and broad aesthetic tendencies are reproduced on social media and which of these gain the highest levels of engagement. This is particularly interesting in terms of the original creations by the series’ fan communities.

In order to address these objectives, we decided to triangulate quantitative and qualitative research methodologies. In fact, we used a series of social media scraping tools to study, first, the reactions and conversations generated on Instagram; and, second, the remediations of the TV series’ content on YouTube. Specifically, to collect visual and textual data we used data extraction tools provided by CrowdTangle for Instagram, as well as PhantomBuster and the Digital methods initiative tools for YouTube (Rieder 2015), and then we analyzed these contents using an approach of qualitative content analysis.

At this stage of our research, we decided to focus on Instagram and YouTube. These platforms are particularly valuable to our analysis for several reasons: first, the ease of posting audiovisual content (therefore repro-

ducing the original media of the programme); second, the social dynamic of user interaction; and third, the platforms' significant use rate and visibility. Indeed, YouTube and Instagram are the first and third most used platforms by Italians according to Blogmeter's annual survey (data cited in Dara 2022). What remains of this essay is dedicated to the results of this analysis. It is divided into two main sections: the first is dedicated to user interactions on Instagram, the second to remediating images on YouTube. In the conclusion, we offer a handful of reflections on the uses and limits of this approach to media reception studies.

Doc on Instagram: Empathy and the Pandemic

Starting with Instagram, we identified the most popular accounts related to *Doc – Nelle tue mani* in terms of followers. From this point of view, the most popular profile is the official account of the TV show; this is followed by other un-official accounts created by users, yet which, on closer analysis, were not considered to be of particular interest in terms of the number of interactions, comments, etc. Moreover, these secondary accounts raised some problems in terms of searchability and permanence of data. For this reason, we decided to focus our analysis, at this stage, on the official TV show account.

In terms of followers, the profile saw a 96% increase between the first and second season (March 2020-March 2022), but there were several growth phases consistent with the 'bumpy' airing of the TV series caused by the spread of the COVID-19 epidemic that impacted the production of the TV show, which had to be stopped due to the lockdown in Italy from 9 March to 3 May 2020. The first season therefore aired after the lockdown had started, i.e. from 26 March to 16 April for a total of 8 episodes, and had to be stopped in the middle because the remaining 8 episodes had not yet been fully filmed. In this first phase, the ratings were record-breaking, reaching almost 9 million viewers.

The growth of followers in this first season was parallel to that of the audience, rising from less than 7 thousand to more than 51 thousand (+87%). Once the lockdown ended, the crew returned to the set and the second part of the first season aired from 15 October to 19 November 2020, with peaks of 8.5 million viewers for the last episode. In this second airing, followers

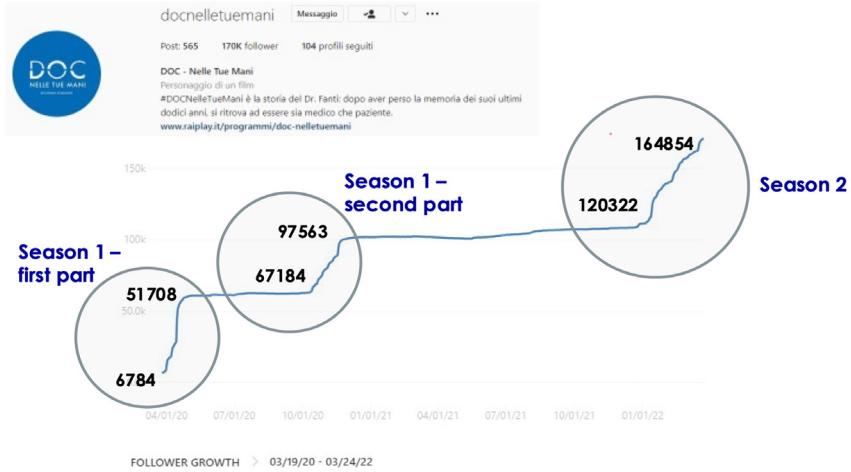


FIGURE 1

Follower growth of the official Instagram profile “Doc – Nelle tue mani” in the period 03/19/20-03/24/22.

increased from 67 thousand to 97 thousand (+31%). The second season aired from 13 January to 17 March 2022 with a peak of 7.3 million viewers. The account’s followers increased in line with the airing, from 120 thousand to 164 thousand (+27%). Each season or partial season thus brought more followers to the show’s official profile and clearly also increased the number of comments and reactions to the posts produced.

From this point of view, therefore, it may be interesting to identify which posts recorded the highest number of interactions in total. The post with the largest number of interactions in the first part of the first season is the one announcing the season finale and which, through a brief synopsis of the episode, invites the audience not to make any plans for the next day, with 36 thousand interactions (Fig. 2).

These increase considerably in the second part of the first season, with a peak of 101 thousand interactions recorded by a video post that captures a central scene of the twelfth episode, in which the protagonist, Dr. Fanti, gives a motivational speech to his team, inviting them to try to do their best for their patients, standing between them and death, because being doctors means not giving up and repeating to oneself the words “non oggi”, “not today” (Fig. 3). The post with the most interactions of the second season registers 235,000 and features various images and a video from the season



FIGURE 2
Top post of first part of the first season.

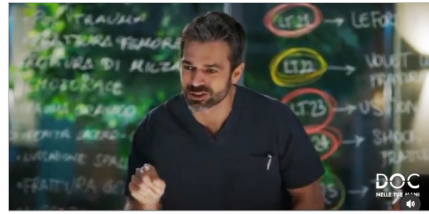


FIGURE 3
Top post of second part of the first season.



FIGURE 4
Top post of the second season.

finale, a very moving topical scene in which the death of one of the team's female doctors is recounted, thus favoring the sentimental plot (Fig. 4).

These quantitative data are worthy, in our opinion, of a more in-depth qualitative analysis that allows us to capture more interesting and subjective aspects regarding viewers' interpretations, attitudes and experiences. The institutional image, featuring the cast of the series as a teaser for the last episode, induces fans on the one hand to express their appreciation for a TV show judged to be of great quality, worthy of appearing on OTT platforms ("Questa serie merita di essere proposta sulla piattaforma Netflix o simili"), and on the other hand to express their regret at leaving so soon a TV programme they had grown fond of ("Non ci posso pensare che domani è l'ultima puntata..."). Some of the comments also question the reason for such an unexpected conclusion ("Ma perchè così poche puntate?"). The comments, at this stage, tell of an audience willing to become attached to this product and to create those parasocial relationships that are so important for loyalty to contemporary seriality (cf. e.g., Mittel 2015, Eyal and Cohen 2010, Lather and Moyer-Guse 2011, Russel and Jensen Schau 2014) and that will be demonstrated in the comments to the following top posts.

The comments on the October 2020 post, dedicated to a topical scene which could be related to the professional plot, besides highlighting the skills of the actors, show the audience high level of self-reflection about the pandemic period that has just been experienced. Most of these, in fact, are thoughts about the pandemic and the role of medical employees, their dedication and sacrifice (“In questo periodo storico non diciamo neanche un grazie ai medici che sono gli unici che non si sono fermati dall’arrivo del covid 19”).

There are also those who report on personal life events, family bereavements (“Doveva essere un non oggi anche per mia madre”), which also testifies to how this TV show has been able to graft itself into the particular historical period in which it aired and, at the same time, into the lives of the viewers, who recognized its value of “mediated authenticity” (cf. Hill 2018), which is one of the keys to audience engagement. Many comments, moreover, note the strong emotional impact of the scene and the emotional realism aroused, mainly due to the actors (“Bravissimo è riduttivo ❤️”; “Stupendo e credibile 🤔👏”). Yet other comments show an intertextual competence of the viewers who recall how the words spoken by Dr. Fanti are reminiscent of those pronounced by Arya Stark *Game of Thrones*: “Not today” becomes “Non oggi” in *Doc*.

The very top post in terms of number of absolute interactions is the one showing one of the most moving scenes of the season finale, which we can relate to the sentimental plot, and obviously account for the great emotions aroused by the scene: “Quanto ho pianto 😭”; “Ho pianto come se fosse una mia parente”; “Lacrime a non finire 🥲🥲”. This aspect is often linked to the skill of the actors and the quality of the TV series, whose realism is particularly appreciated (“Grazie a tutti per l’interpretazione e la credibilità che avete dimostrato ❤️”). Another interesting phenomenon emerging from the comments to this top post lays in the parasocial relationships that have been built up over time with the characters, so that the death of one of them also elicits negative reactions and comments, because the viewers grieve the detachment (“Non dovevate far morire Alba!!!”; “No non mi è proprio piaciuto 🙄”). There are also comments wishing for a third season, and some others demonstrating the audience’s textual competence in comparing this season with the previous one and pointing out its lower quality. It is worth noting the presence, among the latter, of appreciative comments in Spanish, Portuguese, and French, attesting the international circulation of the series, sold in several countries.

If we move on to consider the interaction rate by followers, that is to say interactions divided by number of posts divided by the average followers over the timeframe, the highest rate can be found in a post about the first episode of the first season (34.37% of interaction rate), followed by a post about the last episode of the first season (almost 21% of interaction rate), followed by a post from August 2021, when the series was not airing (almost 21% of interaction rate) and a post about the second episode of the second season. Due to the peculiarities of the comments, we will briefly focus on the last two.

The post of August 2021 is a video in which Doc. Andrea Fanti, invites the Italian population to vaccinate, in order to beat COVID-19 together. The video represents an advocacy activity, on behalf of the TV program and of Italian public service broadcasting more generally, at a time that was still crucial to the fight against Coronavirus. The overwhelming majority of comments express applause, agreement with the message and therefore with vaccination (“Giusto DOC ❤️. Vaccinatevi”; “Bravo DOC”; “Esatto, se no non ne usciremo mai”). Some also recognize the value of the operation of raising awareness of vaccination through a particularly popular character and series (“Bravi @docnелtuemani, anche l’intrattenimento deve sensibilizzare ❤️❤️❤️”). Several ironic comments show willingness to be vaccinated or to return to be vaccinated if the doctor was Doc (that is the actor Luca Argentero portraying Dr Fanti), while only very few comments show opposition to this message and refer to freedom of choice or the lack of transparency of the operation (“Ma la libertà di scelta dove è finita???”). In this sense, therefore, the comments do not express the polarization one would have expected from such a sensitive topic, but seem to almost totally support the case for vaccination.

Another interesting post from our point of view is the one on the second episode of the second season, because in this case a polarization clearly emerges between those who appreciate the show’s restart with its twists and turns, the realism with which the pandemic period and the events experienced by the medical and paramedical staff were recounted (“Bellissimo e straziante allo stesso tempo”; “Un inizio col botto!!!! 😊”). But on the other hand many comments showed impatience with these contents and the need for escapism, for light-heartedness precisely because the experience of the pandemic was not completely processed and still too close (“Non avrei affrontato il tema del COVID-19, siamo già stressati, preoccupati e abbastanza saturi”). In particular, negative comments are also expressed be-

cause of the disagreement with the death of a character, to whom the audience had become attached and whom they do not want to let go; this loss is experienced as another bereavement and its logic in terms of narrative development is not fully understood (“Che senso ha la morte di Lorenzo?? Continuo a non capire e soprattutto non riesco a smettere di pensarci, come vi è venuto in mente??? Incredibile 🤖🤖”).

In short, therefore, we can say that the number of interactions and comments is favored and is closely related to the sentimental plot, but also that medical and professional issues created a significant amount of engagement, especially considering how much the experience of the pandemic and all the emotions aroused by it were concurrent with the airing of the first season and to be fully elaborated in conjunction with the airing of the second season.

YouTube: Top-Down Promotion, Bottom-Up Emotion

The second part of our analysis concentrated on the video-sharing platform YouTube. As mentioned, the objectives of this section were to understand the typology of videos posted with content relating to *Doc*, the modalities through which the series is used on the platform within top-down marketing and bottom-up fan practices, and which of these typologies and modalities gained the highest interest from users (in terms of views, likes and comments).

In order to address these questions, we compiled two datasets, between 30 September and 2 October 2022, each using a variety of automated and manual data scraping:

The top 500 video results from the query “Doc nelle tue mani”. This data was obtained using the YouTube API scraping tools designed by Bernhard Rieder and his team, available via the Digital Methods Initiative (Rieder 2015). These results consist of the top 500 videos geo-localized to Italy, reproduced according to “Relevance” – this is YouTube’s term, as per the advanced search filters – with no personalization (i.e., no personal account data or history influencing the results).

The videos with over 25,000 views that emerge from the query “Doc nelle tue mani” on YouTube – again, with no personalization of the account, geolocalized to Italy – with results ordered by number of views. This results in 93 videos. Further specific information about these videos (title, user,

views, comments, hashtags, etc.), were obtained using Rieder's YouTube tools and the API scraper PhantomBuster.

A necessary note on methodology: whether organized by relevance or by number of visualizations, the results in the datasets are nonetheless subjected to further manipulation by YouTube's algorithm. As scholars have indicated, the platform's recommendation system is reliant on a number of different factors – including previous popularity, publication date and thematic affiliation – but also broader social dynamics such as viewing context and (histories of) individual agency (cf. Airoidi et al. 2016, Rieder et al. 2018). The algorithm is also subject to historical change, since the launch of YouTube (Bärthel 2018). As Arthurs, Drakopoulou and Gandini write (2018: 11), faced with this algorithmic fluidity, “research has to comply with the role of this ‘black box’ algorithm within the specific dynamics observed” (on the concept of the black box, the authors quote Pasquale 2015).

This is the motivation behind our use, for this part of the research, of different datasets: by approaching the content from two different perspectives (relevance and view count/popularity), we hope to guarantee a more representative sample of videos and mediate, to some extent, the subjectivity that the algorithm necessarily enacts. We recognize, however, that the datasets remain a single, personalized snapshot of an otherwise fluid scenario. Indeed, while the results of dataset 2 hypothetically prioritize the most popular videos, we found it still excludes some high-performing videos with over 25,000 views (these were nevertheless excluded from that dataset). Furthermore, while we sought to make the searches as “neutral” as possible, searching without an account (and its viewing history) that would skew results, this nevertheless makes it somewhat unrepresentative of those users – such as specific *Doc* fans – whose own history would indeed skew the results. In short, these videos are not comprehensive, but we consider them to be useful and indicative of general and popular video content regarding *Doc*.

The first dataset (500 videos, “relevance”) indicates an initial set of characteristics of the YouTube videos that engage with *Doc*. First, as illustrated in Figure 5, video uploads are concentrated around the release dates of the episodes of *Doc* on Rai 1. Moreover, if we compare upload date to number of views, as in Figure 6, there is a trend of larger visualizations around release date, indicating more intense interest during the broadcast periods. The figure also demonstrates the growth in interest from season 1 to season 2.

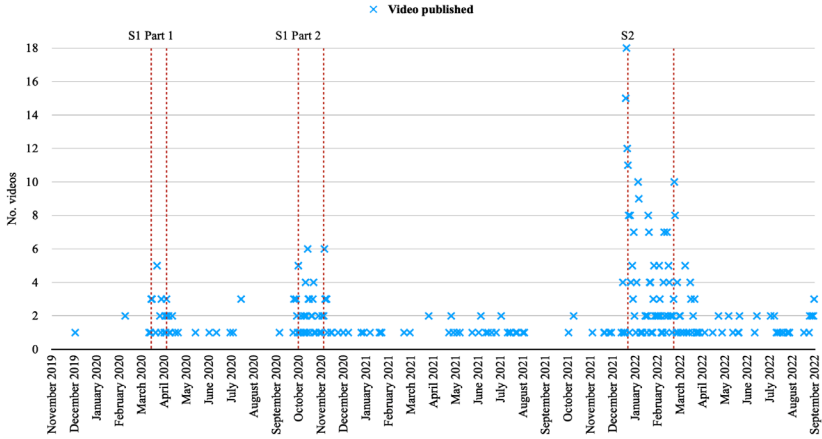


FIGURE 5 Upload dates of YouTube videos and episodes of *Doc* (first dataset, 500 videos, criteria: relevance). The broadcast periods of the two seasons are indicated with red dotted lines.

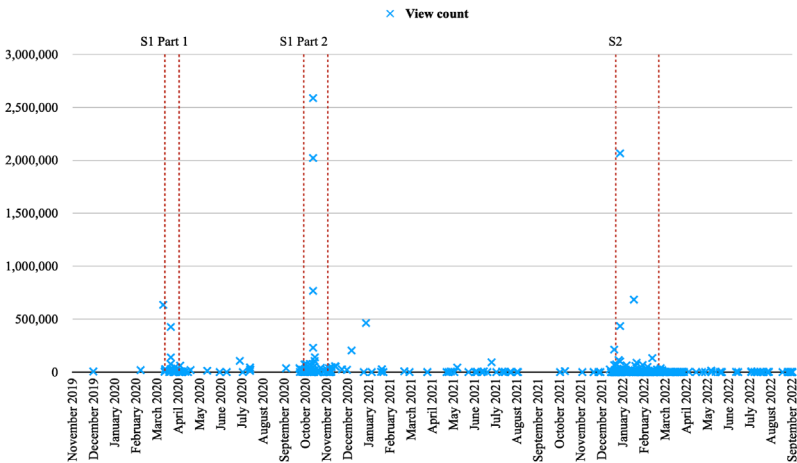


FIGURE 6 View counts of YouTube videos relating to *Doc* (first dataset, 500 videos, criteria: relevance) and release date, with TV broadcast periods indicated in red dotted lines.

The other principal trend that emerges from the first dataset relates to the YouTubers who appear most recurrently in the list of videos that interact with the series. The top ten channels that recur the most are detailed in Table 1.

CHANNEL	N° VIDEOS	VIDEO TYPE	TOTAL VIEWS FOR THE VIDEOS	AVERAGE VIEWS PER VIDEO
daburch86	67	General TV content	13010	194
Beth Patton	34	General TV content	3110	91
Anlove 84	34	Remixes	50234	1477
UD News	21	Recaps	271620	12934
Alexia	17	Recaps	48809	2871
Rai	9	Promotion, Interviews	556289	61810
Various Artists – Topic	9	Music videos	8324450	924939
JupiterOriny1	7	General TV content	8868	1267
spettacolomania.it	7	Entertainment	286424	40918
Aboutpeppe893	6	Influencer (doctor)	90570	15095

TABLE 1

The YouTube channels with the highest n. of videos relating to *Doc* (column 2), also indicating total and average views for the same videos.

The two highest channels (daburch86 and Beth Patton) are anonymous accounts that reproduce high quantities of short videos of Italian television content. The columns referring to total video views per channel, and particularly the average video views, signal that their content is spread very thinly. We can hypothesize that this is a larger strategy of minimizing production costs and mass-producing content in order to maximize the possibility of some videos gaining high views and subsequent ad profits. Indeed, at time of writing, only 25 of the more than 20,000 videos uploaded on the daburch86 channel have more than 10,000 views (the average is 130 views). This appears to indicate a twofold dynamic: on the one hand, the

potential for the audience's interest in *Doc* within a much broader category of remediated TV content on YouTube; on the other, an awareness of the possibility to monetize that content, including *Doc*, within a “grey zone” – the term is Lobato's (2012: 95) – with regard to copyright law.

The ways that *Doc* content is integrated within YouTube's industrial dynamics around macro themes (TV, music, but also medical content broadly) is also made evident by a pair of videos uploaded by the YouTuber “Aboutpepe893”. The user is a doctor who posts content on his own experiences and ironic responses to popular culture. The videos on *Doc* – reaction videos and evaluations of the medical science, all comical in nature – are quite critical of the series. Evidently *Doc* informs just a small fraction of his overall output (six of 811 videos). This content also has a slightly lower average view count (c. 15,000) than for his channel overall, (c. 21,000 per video, though a couple have over 1 million). As with the previous examples, then, *Doc* is integrated into a larger content tendency, and therefore visible, we might assume, to users that are not necessarily fans. The fact that Aboutpepe893 has a relatively high rate of views and average views per video – at least, among the channels in Table 1 – is further testament to how the show can be adapted to broader “grammars” of social media (as per the “Jerusalem” trend, to which we return below). Indeed, not by chance he often reminds viewers to “like and subscribe” – well-aware that it is socially and economically beneficial in terms of visibility.

The most popular individual videos are more comprehensively indicated by our second dataset (criteria: popularity). When recording data for the 93 videos with the highest view counts, we also manually added a typology of the video's genre. This led to seven categories: (i) music videos; (ii) interviews with the cast; (iii) promotion; (iv) recaps; (v) remixes of content from the show; (vi) reproduced scenes; and (vii) miscellaneous: skits, news, behind the scenes, or commentaries.

The most common categories per number of videos are remixes (33% of the videos), music videos (27%) and interviews with the cast (21%); the other categories make up the remaining 19%. “Remixes” are videos of *Doc* content that are re-edited, with a musical soundtrack played over the top. The vast majority of these are dedicated to the romantic relationships between the staff at the hospital – Giulia and Lorenzo, Giulia and Andrea, Andrea and Agnese, Alba and Riccardo. There are some other more broadly thematic videos, referring to, for example, the “not today/non oggi” sentiment that also emerges powerfully in Instagram content, as mentioned above.

The second category is the music videos of the *Doc* soundtrack. The most viewed videos are the topic videos, consisting of a still image of the series and one of the songs; some other videos also contain footage; others still are popular covers of the songs.

Most striking about these videos is the fact that, by a significant amount, they are the most popular videos in terms of views. Indeed, if we take the total views in this sample and re-consider our 7 categories as *shares* of those views, rather than total number of videos, the music videos represent 67% of the views. Three of the 25 videos in this category have over 2 million views (the songs “Changed at All”, “Firestones (feat. Nico Bruno)” and “I’ll Find You in the Dark”, all composed by Tony Brundo); the average view count is 500,000.

The third category is videos reproducing whole interviews, or segments of them, with a variety of the cast members. Most predominant among these is, unsurprisingly, Luca Argentero (a prominent film and television star in Italy who plays the part of Dr. Fanti); we also find many interviews with Matilde Gioli, the actress who plays Giulia, and, to a lesser degree the other actors/doctors. There is a prevalence of cast-members who are also recognizable outside of this programme (e.g., Giovanni Scifoni, Pierpaolo Spollon, Gianmarco Saurino).

Beyond these broader tendencies, we can also identify a handful of other particularly popular videos that signal relative tendencies of the show’s reception. The first of these is the single example of a clip from within the series: the characters dancing to “Jerusalem”, by Master KG, after a grueling day fighting COVID-19 in episode 2x08 (“Cane blu”). Though part of the episode, the clip is essentially the reproduction of a dance challenge that was popular on social media between 2020 and 2021, and hence the popularity of the video can also be understood within that dynamic. At the time of the study, the video had 683,000 views, 7,000 likes and over 300 comments – the most of any video across the samples. Space limitations prevent a detailed analysis of these – this also lies beyond our objectives for the analysis of YouTube. Limiting ourselves to a superficial overview, however, it is worth noting that the user response to this video is broadly in line with the responses on Instagram, too. Of the 311 comments recorded at the time of the study,³ the vast majority of them are simple, positive responses

³ All comments were recorded on 30-09-2022 from the following video: “Doc 2, Ar-

to the scene, the music, the stars and/or to the show more generally: “Adoro questa scena!!!!”; “I medici che tutti vorremmo”; “Serie televisiva fantastica adoro tutti in particolare Argentero 😊😊😊”. Many of these contain explicit references to emotional responses to the scene: “Mi sono commossa, solo leggere la data 2020, mi è venuta la pelle d’oca.”; “Senza parole ...solo grandi emozioni ❤️”. There are a handful of more negative responses – a few denounce the quality of the show and of Italian TV fiction generally, and quite a few make references to the satirical TV show *Boris* (2007-10, 2022), either denouncing this as the confirmation that *Boris* foresaw a (terrible) future (“Viviamo in una distopia e boris aveva ragione”) or simply commenting famous, satirical lines of dialogue (“la locura René”; “Un paese di musichette, mentre fuori c’è la morte”).

As the latter point anticipates, current affairs are present on YouTube, too, including a handful of lengthy diatribes about the invasion of Ukraine and “World War III”, and, more importantly, references to COVID-19 and the Italian management of the pandemic. The majority of the latter are positive (“Un grazie con tutto il cuore a tutti i medici , operatori sanitari, tutto il personale medico, ecc. che ha combattuto e combatte il covid. 💕💕💕”); “Bella puntata dopo il covid ci voleva un bel balletto bravi A tutto il cast 🎯🎯”), though some others are more skeptical, arriving even into the realm of conspiracy theories (“IL LIVELLO DI DEMENZA DEGLI ITALIANI È GARANZIA DI SUCCESSO PER QUALSIASI DITTATURA , NON A CASO ABBIAMO INVENTATO ED ESPORTATO IL FASCISMO OVVERO LA DITTATURA DEL CONFORMISMO AVANTI CON LE DOSI E OFFRITE IL BRACCIO ALLA PATRIA VACCINO E MASCHERINA VERSO IL SOL DELL’AVVENIRE”). None of the comments reproduce or cite Dr Fanti’s “non oggi” speech, however. Finally, once again, there are also a number of comments in French, Spanish and Portuguese, which further demonstrate the arrival of the show in global markets.

Relatedly, another high performing video, from the category of promotional videos, is the trailer for *Doc – Nelle tue mani* for Amazon Prime Brazil. The video has a high interaction rate: at the time of the study it had over 90,000 views, making it the 22nd most popular video, though in terms

genero e il cast ballano sulle note di Gerusalemme”, *Quotidiano Nazionale*, 8 February 2022, <https://www.quotidiano.net/magazine/video/doc-2-argentero-e-il-cast-ballano-sulle-note-di-gerusalemme-1.7341449>. The spelling, punctuation and emphases are reproduced sic, as per in the comments, with no further editing.

of comments – it has 200 – it is the third most popular video in the sample. The comments are almost entirely in Portuguese, with very few exceptions (in Spanish), and they are overwhelmingly positive. In fact, it is difficult to find any negative comments at all, with the exception of one complaint that the trailer presents too much of the story, and a handful that bemoan the lack of an original language audio track on Amazon Prime Video. The positive comments once again pick up on the humanity and sensibility of the show, the positive performances of the actors, and the soundtrack. Of course, it is possible that Amazon Prime Brazil simply repressed any negative comments, with an eye to positive promotion. The outlying popularity and interaction rate remain nonetheless an indicator of the series' success in the Brazilian market.

The nature of the most popular videos that include *Doc* content provides the stimulus for some interesting reflections. First, the prevalence of music is striking, revealing a broader impact of the series that transcends to another media: social media and pop music. This also finds confirmation in the second most common category, that of the remixes, which often also include songs from the soundtrack (especially “Changed at All”), as well as in the comments of two videos mentioned here. Secondly, the high percentage of remix videos signals another interpretative key for this series, that evidently foregrounds the “sentimental plot” – one of the three key “isotopies” within the medical drama identified by Pescatore and Rocchi, that emerges horizontally across seasons. These videos indicate the high level of its resonance with fans – especially since the amnesia trope used in *Doc*, as well as some characters deaths, provides the potential to promote different couples at different times.

The medical plot isotopy, on the other hand, is somewhat pushed to the side. This does not appear in any of the highest performing videos relating to *Doc*, with the exception of those produced by real-life doctor and influencer Aboutpepe893. It is implicit in the Jerusalema video, and prevalent in its comments, though there is little indication that we would find the same results in the comments more broadly. Indeed, in the comments on the Prime Video Brazil trailer, COVID-19 is rarely mentioned, and only in anticipations of the second season which, at that time, had not been released in the Prime Video catalogue in that country.

The final and most manifest indication that emerges from the YouTube datasets is that *Doc* exists not as a single phenomenon or a unique product of entertainment or art, but rather the show is adopted and adapted in fluc-

tuating media-industrial dynamics, with aims that shift from promotion of the series to fan appreciations of certain plots, and from monetizing macro themes (medicine, Italian TV) to emotional reactions to songs.

Conclusions

There is evidently much more that could be said about the presence of this series on social media, in relation to the small amounts of data that we have already collected and, of course, by expanding and correlating it with other trends. For now, though, suffice it to indicate some preliminary conclusions. The analysis of comments and engagements on Instagram and YouTube help us to situate the ways that users and fans interact with and respond to *Doc*, therefore demonstrating how social media are a useful arena – alongside the media product itself and ethnographic study – through which it is possible to retrace reception trends.

Our analysis shows that social media engagement is very closely tied to traditional broadcasting, and that the popularity of Rai content on both platforms indicates the relevance of promotional materials as anchors for this engagement. User comments and video uploads reveal the centrality of the sentimental plots of the show, though the medical isotopy also emerges in the medical science (or lack thereof) in *Doc*, and especially through the show's connections to the pandemic. The few examples considered here also signal the multimedia nature of TV consumption here in Italy, as the show itself is tied to other TV productions (in comments, but also as content for high-output generalist TV YouTube accounts), to social media creations, and to the consumption of popular music.

A further trend that emerges quite manifestly from the analysis of Instagram and YouTube is the constant presence of highly impassioned content, through emotionally-charged language and emojis in the comments (both in response to the show and the stars, but also in relation to the bigger global issues of the invasion of Ukraine and the pandemic) but also and especially through fan-made content such as remixes. This ties, broadly, to a characteristic of the current conjuncture regarding the role of affect in popular culture, and more specifically to the attention economy on social media – since affect can be a useful tool to promote high visibility. In their article on YouTube as a research tool, Arthurs, Drakopoulou and Gandini note this tendency, referring repeatedly to a case where “a YouTuber’s success and the

outcomes one is able to elicit are partly dependent on the capacity of a certain video to ‘work’ the platform’s infrastructure to an extent that the processes of affect-based monetization are enabled” (Arthurs et al. 2018: 10). Drawing from this context, we hypothesize that the pairing of “affect-based monetization” could be a useful interpretative key for reception studies on social media – especially in the context of the medical drama, considering the same prominence of the sentimental plot.

The analytical methods used in our reception analysis also face some challenges and limitations, of course. The first of these relates to the difficulty of gaining the “bigger picture”, both due to the social media algorithm’s manipulation of search results, but also to the difficulty of finding content that flies below the radar, i.e., that does not emerge in search results as it does not explicitly reference the series. The second issue instead relates to the heterogeneity of interactions: simply viewing, liking, and commenting are three very different actions based on a great variety of responses. The comments that we see here, as we have illustrated, appear to represent positive or negative positions; the middle-ground remains unclear. With these limitations in mind, it is clear that this initial investigation, while confirming a set of useful tendencies, is a partial view that necessitates further research and expansion with findings from other methodologies and critical-theoretical standpoints (affect theory might be productive, for instance). It is our hope that we will be able to address some of these issues and compensate for the limitations they pose by combining these findings with other methods throughout the *NEAD* project.

BIBLIOGRAPHY

- Airoldi, Massimo, Davide Beraldo and Alessandro Gandini (2016). "Follow the Algorithm: An Exploratory Investigation of Music on YouTube." *Poetics* 57: 1-13. <https://doi.org/10.1016/j.poetic.2016.05.001>.
- Arthurs, Jane, Sophia Drakopoulou and Alessandro Gandini (2018). "Researching YouTube." *Convergence: The International Journal of Research into New Media Technologies* 21(1): 3-15. <https://doi.org/10.1177/1354856517737222>.
- Bärtl, Mathias (2018). "YouTube Channels, Uploads and Views: A Statistical Analysis of the Past 10 Years." *Convergence: The International Journal of Research into New Media Technologies* 24(1): 16-32. <https://doi.org/10.1177/1354856517736979>.
- Brembilla, Paola and Ilaria A. De Pascalis (eds) (2020). *Reading Contemporary Serial Television Universes: A Narrative Ecosystem Framework*. New York-London: Routledge.
- Dara, Virginia (2022). "Italiani e social media 2022: i dati di Blogmeter." *Inside Marketing*. <https://www.insidemarketing.it/italiani-e-social-media-2022-dati-blogmeter/>.
- Ducheneaut, Nicolas, Robert J. Moore, Lora Oehlberg, James D. Thornton and Erik Nickell (2008). "Social TV: Designing for Distributed, Sociable Television Viewing." *International Journal of Human-Computer Interaction* 24(2): 136-54. <https://doi.org/10.1080/10447310701821426>.
- Eyal, Keren and Jonathan Cohen (2010). "When Good Friends Say Goodbye: A Parasocial Breakup Study." *Journal of Broadcasting & Electronic Media*. 50(3): 502-523. http://dx.doi.org/10.1207/s15506878jobem5003_9.
- Fumarola, Silvia (2020). "Ascolti 'Doc', Argentero torna e sbanca l'Auditel, quasi 7 milioni e mezzo per la serie di Rai 1." *La Repubblica*, October 16. https://www.repubblica.it/serietv/rai/2020/10/16/news/doc_ascolti_nuovi_episodi-270688587/.
- Hill, Annette (2018). *Media Experiences: Engaging with Drama and Reality Television*. New York-London: Routledge.
- Lather, Julie and Emily Moyer-Guse (2011). "How Do We React When Our Favorite Characters Are Taken Away? An Examination of a Temporary Parasocial Breakup." *Mass Communication and Society* 14(2): 196-215. <http://dx.doi.org/10.1080/15205431003668603>.
- Levine, Elena (2001). "Toward a Paradigm for Media Production Research: Behind the Scenes at General Hospital." *Critical Studies in Media Communication* 18(1): 66-82. <https://doi.org/10.1080/15295030109367124>.

- Livingstone, Sonia (2013). "The Participation Paradigm in Audience Research." *The Communication Review* 16(1-2): 21-30. <https://doi.org/10.1080/10714421.2013.757174>.
- Lobato, Ramon (2012). *Shadow Economies of Cinema: Mapping Informal Film Distribution*. London: Palgrave Macmillan-BFI.
- Mittel, Jason (2015). *Complex TV: The Poetics of Contemporary Television Storytelling*. New York: New York University Press.
- Nicholas (2022). "Doc – Nelle tue mani 2: analisi ascolti della seconda stagione." *BubinoBlog*. <https://bubinoBlog.altavista.org/doc-analisi-ascolti-della-seconda-stagione-della-serie-tv-con-luca-argentero/>.
- Pasquale, Frank (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.
- Pescatore, Guglielmo (ed.) (2018). *Ecosistemi narrativi. Dal fumetto alle serie TV*. Roma: Carocci.
- Pescatore, Guglielmo and Marta Rocchi (2019). "Narration in Medical Dramas I. Interpretative Hypotheses and Research Perspectives." *La valle dell'Eden* 34: 107-15.
- Ranjana, Das and Brita Ytre-Arne (eds) (2018). *The Future of Audiences: A Foresight of Interfaces and Engagement*. Cham: Palgrave Macmillan.
- Rieder, Bernhard (2015). "Home." YouTube Data Tools (Version 1.30) [Software]. <https://tools.digitalmethods.net/netvizz/youtube/>.
- Rieder, Bernhard, Ariadna Matamoros-Fernández and Òscar Coromina (2018). "From Ranking Algorithms to 'Ranking Cultures': Investigating the Modulation of Visibility in YouTube Search Results." *Convergence: The International Journal of Research into New Media Technologies* 24(1): 50-68. <https://doi.org/10.1177/1354856517736982>.
- Rocchi, Marta (2019). "History, Analysis and Anthropology of Medical Dramas: A Literature Review." *Cinergie* 15: 69-84. <https://doi.org/10.6092/issn.2280-9481/8982>.
- Russell, Antonia and Hope Jensen Schau (2014). "When Narrative Brands End: The Impact of Narrative Closure and Consumption Sociality on Loss Accommodation." *Journal of Consumer Research* 4: 1039-62. <https://doi.org/10.1086/673959>.
- Schröder, Kim Christian (2019). "Audience Reception Research in a Post-Broadcasting Digital Age." *Television & New Media* 20(2): 155-69. <https://doi.org/10.1177/1527476418811114>.

TV RECEPTION VIA SOCIAL MEDIA ANALYSIS:
THE CASE OF DOC – NELLE TUE MANI



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.6bec3626>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.

CONTRIBUTORS

Stefania Antonioni is associate professor of Television Studies at the Department of Communication, Humanities and International Studies of the University of Urbino Carlo Bo. Her research interests include transmedia narratives, audience reception online and offline, TV serial narratives, visual studies. She authored the book *Imagining. Serialità, narrazioni cinematografiche e fotografia nella pubblicità contemporanea* (2016) and several articles, among which “SKAM Italia Did it Again’: The Multiple Lives of a Format Adaptation from Production to Audience Experience”, in *Critical Studies in Television* 16(4) 2021 (with L. Barra and C. Checcaglini); “From Parenthood to *Tutto può succedere*: Ready-made elements and cultural translation”, in *VIEW Journal of European Television History & Culture* 9(17) 2020 (with C. Checcaglini).

Giorgio Avezzi is assistant professor at the University of Bergamo (Italy). He is the author of two books dedicated to the relations between audiovisual content and geography; the most recent one (*L'Italia che guarda*, 2022) discusses the geographies of audiovisual consumption in Italy. He is also co-editor of two journal issues dedicated to the relations between geography, cinema, media and visual culture, *Cinergie* 10 2016 and *Necus: European Journal of Media Studies* 7(2) 2018. Among his most recent publications is an article for *GeoJournal* (2022) mapping the market for foreign cinema in Italy during the last twenty years.

Rosa Barotsi is a researcher and lecturer at the University of Modena and Reggio Emilia, Italy. She recently concluded a Marie Skłodowska-Curie Action at the Catholic University of Milan with the project *CineAF: Women's Films in Italy (1964-2015)* on gender inequality in the Italian film industry. She received her PhD at the University of Cambridge in 2014.

She has previously held a postdoctoral position at the ICI Berlin Institute of Cultural Inquiry, where she developed a project on Slow cinema and debt. Along with Clio Nicastro and Saima Akhtar, she co-founded the *In Front of the Factory* research collective in 2016. She is one of the coordinators of the Eurimages-funded joint project *The Purple Meridians*. Her research and curatorial work focuses on the intersections between film, gender and work, with an emphasis on Italian and Greek cinema.

Marta Boni is an associate professor in the Department of Art History and Film Studies at the University of Montreal. Her work covers theoretical and methodological problems linked to serial phenomena and their transformations in the contemporary panorama of digital cultures: TV series, transmedia, worldbuilding, and queer television. Along with various articles and book chapters on contemporary TV, she published *Romanzo Criminale: Transmedia and Beyond* (2013), edited *World Building: Transmedia, Fans, Industries* (2017), *Formes et plateformes de la télévision à l'ère du numérique* (2020), *Intervalles sériels* (with T. Carrier-Lafleur and F. Khazoom, 2021), and *Perdre Pied. Le principe d'incertitude dans les séries* (forthcoming 2023).

Paola Dalla Torre is associate professor of Cinema, Photography, Television at LUMSA of Rome. Her research activity has developed along some guidelines: the study of contemporary cinema in its ethical-philosophical implications; the analysis of the genre of science fiction in the contemporary world; and recently the analysis of TV series through a quantitative methodology. In addition, her research is now focused on the economic-cultural study of Italian film exhibition, within nationally funded research on Italian cinemas with three Italian universities (*Cin_Ex*, PI Mariagrazia Fanchi).

Mirko Degli Esposti is full professor in Mathematical Physics at the University of Bologna. Degree in Physics (Bologna, Italy) and PhD in Mathematics (Caltech and PennState, USA). Postdoctoral fellow at the Mathematical Science Research Institute (M.S.R.I), Berkeley. Visiting Professor at the Georgia Institute of Technology. He has been Head of the Math Department of the University of Bologna, and he served as the deputy Rector of the University of Bologna. Recently he has been visiting Professor at the McGill University (Montreal, Canada). He loves mathematical models of information. In his spare time he bikes. More @ <https://www.unibo.it/sito/web/mirko.degliesposti/en>.

Mariagrazia Fanchi is full professor in Film and Media Studies at the Università Cattolica del Sacro Cuore of Milan, where she is also head of the Postgraduate School of Media, Communication and Arts. Her field of research is focused on cinema and media audiences, gender, and cultural and creative industries analyzed in historical and synchronic perspectives. Since 2019, she has been directing the impact assessment of public funding for cinema and audiovisuals on behalf of the Italian Ministry of Culture. Among her most recent publications are: “Has Digital Transformation Impacted Gender Imbalance in Italian Cinema? A Data Analysis of Creative Clusters 2004-2016” (with M. Tarantino), “Feminine Spectrum: Notes for a History of Italian Women at the Movies” (2019), “For Many but Not for All: Italian Film History and the Circumstantial Value of Audience Studies” (2019), “Donne e reti creative nel cinema italiano (2004-2016)” (2019).

Paolo Fantozzi has a Master’s Degree in Engineering in Computer Science and a PhD in Computer Engineering, both obtained at Sapienza University of Rome. He is an assistant professor at LUMSA University in Rome, where he is in charge of the courses Algorithms and Data Structures, Databases and Big Data, and Data and Social Network Analysis. His research interests are machine learning, artificial intelligence, natural language processing, efficient algorithms on graphs and hypergraphs and theoretical computer science. His Erdos number is 4. He worked on many different research and development projects in many fields: agriculture, transportation, energy management, predictive maintenance, health care, disease monitoring, information retrieval, cyber security, law and regulations. He led the team who created the Voice Assistant at IBM IoT Lab in Hursley in 2018. He has been a tutor at the Italian Olympiads in Informatics since 2018.

Dominic Holdaway is associate professor of Film at the Department of Communication, Humanities and International Studies of the University of Urbino Carlo Bo, where he teaches undergraduate and postgraduate courses in film history, film distribution and media industry studies. His research is focused on the politics of audiovisual media, both in terms of production/distribution politics and representation agendas. He is currently working on a NETIAS CAT project entitled *Screening European Populisms*. He has published various articles on Italian film and television, including “The Neglected Spaces of Feminism and Queer in Contemporary Italian Political Cinema”, in *Zapruder World* (2020) and “*Peppino Impastato* in Made and

Unmade Scripts: The Standards and the Limits of Engaged Cinema”, in *La valle dell’Eden* (2022).

Beata Królicka is a PhD student at Nicolaus Copernicus University in Toruń (Poland). Her fields of interest are media discourse, political communication, sociology of emotions and postcolonial theory.

Andrzej Meler has been affiliated with the Faculty of Philosophy and Social Sciences at Nicolaus Copernicus University in Toruń as a freelancer, researcher, and academic teacher since 2015. Previously, for 10 years, he worked as a research and analysis specialist at the Polska Press Group, where he was responsible for web analytics for news and classified portals. From 2018 to 2022, he served as the main statistician for the PayTech Impact. EU project (<http://paytech.umk.pl>). He has participated in dozens of research projects as a project manager, researcher, analyst, and data explorer. Currently, his research focuses mainly on media (TV, radio, social media) in terms of their pluralism and the role of emotions in media communication. He primarily uses quantitative methods, statistical analysis, and data modelling in his work.

Maurizio Naldi is a full professor of Computer Science at LUMSA University in Rome. He got his PhD in Telecommunications Engineering from the University of Rome Tor Vergata and his MSc in Electronic Engineering from the University of Palermo in 1988. Prior to his academic career, he pursued an industrial career in Several ICT companies, ending as Head, Traffic Forecasting and Cost Analysis in WIND Telecomunicazioni in 2000. He was with the University of Tor Vergata, first as an assistant professor and then as an associate professor from 2000 to 2019. He is Co-Editor of the *Electronic Commerce Research and Applications* journal, published by Elsevier. His research interests span the fields of network and service economics, e-commerce, risk analysis, and applications of machine learning.

Placido Pellegriti received his BSc degree magna cum laude in Computer Science for Data Management from LUMSA University in Rome in 2022 after a six-month internship in Almagiva as an apprentice. Since then, he has been with Almagiva as a Data Scientist. His research interests lie in machine learning and its applications.

Guglielmo Pescatore is full professor of Film and Media Studies at the University of Bologna, where he teaches Entertainment Cultures and Media Economics. His studies on television series gave rise to a strand of research dedicated to narrative ecosystems and vast narratives. Characterized by an interdisciplinary approach, these investigations use qualitative and quantitative analytical tools, some of which are biologically derived. On these topics he published “Narrative Ecosystems. A Multidisciplinary Approach to Media Worlds” (with V. Innocenti, 2017), “The Evolution of Characters in TV Series: Morphology, Selection, and Remarkable Cases in Narrative Ecosystem” (2018 with V. Innocenti), *Ecosistemi narrativi* (2018), “Narration in Medical Dramas I: Interpretative Hypotheses and Research Perspectives” (2019 with M. Rocchi), “Modeling Narrative Features in TV Series: Coding and Clustering Analysis” (2022 with M. Rocchi).

Marta Rocchi is a junior lecturer at the Department of Arts at the University of Bologna. Her research interests concern data-driven approaches for the study of narrative ecosystems, gender inequalities in the audiovisual industry and climate change communications in screen industry. Her publications include a book, several papers in peer-reviewed journals and a book chapter. She was curator of special section of *SERIES* journal on “Making Models of Contemporary Serial Media Products”. Among her most recent publications are: “Modeling Narrative Features in TV Series: Coding and Clustering Analysis” (2022 with G. Pescatore), “Environmental Misinformation and Audiovisual Serial Narratives: An Automatic Analysis of the Twitter Social Discursiveness on *Seaspiracy*” (2022), “Women’s Labour in TV Series Production: A Comparative Analysis of Italian Generalist TV and Pay Platforms (2016-2019)” (2023).

Radosław Sojak is a PhD, post-doctoral degree, assistant professor at the Institute of Sociology at Nicolaus Copernicus University in Toruń, Poland. Junior Fulbright Research Scholar at University of Virginia, Charlottesville (1999). Grantee of Ministry of Science and Higher Education and British Commonwealth Office (Plater/Nuffield Collage at Oxford 1998) and Foundation for Polish Science (twice 2000 and 2003). Author and co-author of five monographs: *The Anthropological Paradox: Sociology of Knowledge as a Perspective In Sociological Theory* (2004); *The Lost Reality: On the Social Construction of the Unknown* (2005 with D. Wicenty); *The Promise Fulfilled: The Preliminary Theory of ‘Research Styles’ – the Case of The Sociology of*

Knowledge (2014); *Aristocrats and Craftsmen: A Synergy of Research Styles* (2015 with Ł. Afeltowicz) and *Citizens Platform's State: A Closing Report* (2015 with A. Zybortowicz and M. Gurtowski). Main fields of interest include: social theory, sociology of science and knowledge, post-communist transformation, public discourse analysis.

Matteo Tarantino is assistant professor at the Università Cattolica del Sacro Cuore of Milan, where he lectures on the sociology of digital technologies. His research interests focus on the intersection between data, society, and culture. Among his most recent publications are: “Navigating the Green Data-scape: Some Challenges in Automating Environmental Data Procurement for Disclosure Efforts in China” (2019), “Uncertainty in the Air: Communicating Urban Air Pollution” (2019), and *Il tecno dragone. L'immaginario tecnologico cinese ieri, oggi e domani* (2020).



